

Conflicts at the NIH (cont.)

Previous assurances by the director of the US National Institutes of Health to Congress over the regulation of conflicts of interest are contradicted by fresh allegations. Tough new rules for staff seem essential to restore public confidence.

Last week, an ongoing investigation into conflicts of interest at the National Institutes of Health (NIH) took a new and more serious turn. Congressional investigators claimed that the NIH had seriously underestimated the number of consulting agreements that its scientists held with outside entities. After contacting pharmaceutical companies, the investigators claimed to have found about 100 instances in which NIH scientists had consulted with industrial partners without notifying the agency.

The investigations subcommittee of the House Committee on Energy and Commerce has yet to look into all these cases, but a few examples cited at the hearing seem to undermine what NIH director Elias Zerhouni has previously stated: that NIH employees have always rigorously adhered to ethics rules. Now it seems that, at least in some cases, NIH employees circumvented the regulations that require them to ask for approval for their paid consulting activities. The committee is alleging that there is more than just smoke — there is fire too.

As yet there has been no response from the employees who have been singled out in the congressional hearing. They face several specific accusations, including misleading the committee in previous testimony over the extent of consultancies with biotechnology companies. The NIH says that some of the discrepancies between its own accounting and that of the committee are due to mistakes in its own paperwork, rather than wrong-doing by its employees. It also says that, in some cases, discrepancies occurred because the NIH tallied consulting arrangements differently from the pharmaceutical companies.

Zerhouni has now told the subcommittee — chaired by James Greenwood (Republican, Pennsylvania) — that “drastic changes” must

occur at the NIH. He has proposed a ban on paid consulting work by senior NIH officials. Less senior employees could still consult, but would only be able to receive the equivalent of 25% of their salaries in consulting payments, and must limit their consulting time to 400 hours. No NIH employee would be allowed to accept stocks as payments. And many scientists would be prevented from holding any stock at all in a drug or biotechnology company.

Some of these and other proposed changes go far beyond those recommended by a specially commissioned panel that issued its report in May. And many NIH scientists are disturbed by them. Some have vowed to leave the agency if the changes require them, for instance, to sell long-held stocks in drug companies. But if the problems are as significant as they appear, the NIH must make some sacrifices now if it is to restore its reputation.

The rules on consultancies were loosened in 1995 under its director at the time, Harold Varmus. Shortly afterwards, the NIH benefited from a lobbying campaign that resulted in a doubling of its budget between 1998 and 2003. The new allegations raise questions about whether the NIH’s leaders during that time — Varmus followed by acting director Ruth Kirschstein, who led the agency from 2000 to early 2002 — did all that they could to ensure that their house was in order, when a funding boom was bound to bring closer scrutiny.

Congress being what it is, there is more than a whiff of political opportunism in the handling of these allegations. But long-standing advocates on behalf of the NIH say that the conflict-of-interest scandal has crippled their ability to lobby for funds. Zerhouni’s proposed restrictions on his colleagues may be drastic, but they may also be exactly what the NIH needs if it is to regain its status as the crown jewel of the federal government. ■

Risks of high winds

Proponents of turbines on top of New York’s Freedom Tower had better get their sums right.

Of all the possible sources of renewable energy, wind farms have provoked some of the strongest resistance. Objectors point to a variety of concerns, from the risk posed to migratory birds to the chance that the spinning blades could fling ice in all directions during winter. But most of all, it seems, people just don’t want the towering turbines to spoil the views from their back gardens.

So proponents of wind power were thrilled to learn in December that the current design for what is to be the world’s tallest building — to be erected at the site of the terrorist attacks on 11 September 2001 in New York — calls for a massive wind farm in its upper storeys. When completed, the Freedom Tower and its turbines will be visible to millions (see page 12).

Provided, that is, that the wind turbines are ever installed. They were the brainchild of the architects Skidmore, Owings & Merrill. But people who build wind turbines for a living say they doubt whether the architects consulted wind engineers before floating the proposal. They say the structural challenges involved in damping

vibrations and noise are too great to be cost-effective. One told *Nature* he thought the proposal was a joke. Another said he didn’t believe that the turbines would ever be built. Skidmore, Owings & Merrill declined *Nature*’s request for a response to the criticism.

Nevertheless, several optimistic companies have submitted proposals. One says it can make the job cost-effective with a turbine that spins on a vertical axis, rather than the usual propeller-like blades that spin on a horizontal axis. This would be good news, but the fact remains that vertical-turbine technology lags far behind and has not yet been used in any successful commercial installation.

Much more is at stake here than just an engineering dispute over what is possible. The amount of electricity involved is small compared with New York’s voracious appetite, but the symbolism is huge. Wind power, whether urban or rural, could help ease us into our inevitable oil-free future. Yet a failure on the scale and visibility of the Freedom Tower could dampen enthusiasm for this much-needed renewable resource far more than a few dead birds could. Let’s hope the optimists have it right. ■





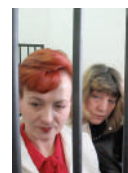
Kerry tunes in
Democrat candidate
pledges to sell radio
spectrum for science
p4



Growing pains
Academics call for
scientific response
to world hunger
p5



Paint plot
University redraws
the boundaries for
illustration course
p6



Freedom bid
AIDS researchers
petition Libya over
condemned medics
p8

Earth science loses autonomy as NASA switches focus to the Moon

Tony Reichhardt, Washington

NASA is poised to undertake its most sweeping reorganization in more than a decade. The plan, announced on 24 June, will see the agency halve the number of offices at its headquarters, eliminate its Earth sciences directorate, and transfer several key personnel, including head of space science Ed Weiler, into new positions.

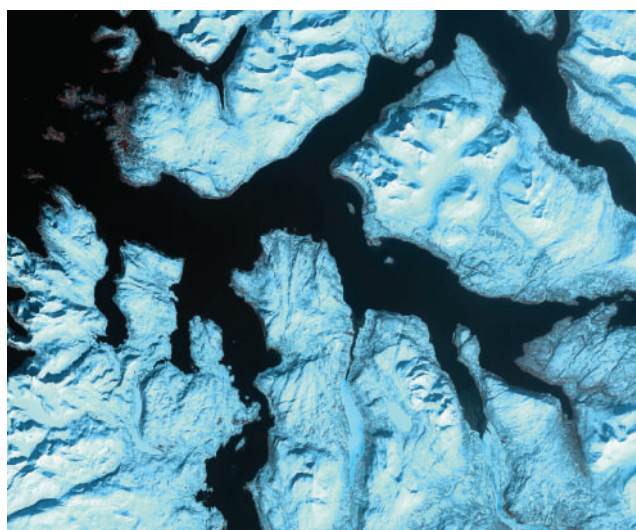
Sean O'Keefe, the agency's administrator, says that the streamlining will focus NASA more clearly on the Bush administration's priorities — particularly sending astronauts back to the Moon and on to Mars. But critics worry that the changes may marginalize the Earth sciences and reduce NASA's involvement in the politically contentious sphere of climate change.

From 1 August, NASA will be organized around four missions: exploration (the new Moon–Mars programme), space operations (including shuttle launches), science and aeronautics. Earth science will be lumped in with space science, but most of the microgravity research done on the International Space Station will fall under the exploration office.

In addition, the agency's ten field centres will report directly to a single mission directorate, rather than serving many masters as they do now. For example, three of the centres — Ames in Moffett Field, California; Goddard in Greenbelt, Maryland; and the Jet Propulsion Laboratory in Pasadena, California — will fall under the science mission.

In merging the Earth and space science programmes, O'Keefe undoes the work of his predecessor Dan Goldin, an appointee of the first president George Bush. Goldin dismantled the larger Office of Space Science and Applications in 1992 and elevated Earth science to an equal footing with space science.

Charles Kennel, director of the Scripps Institution of Oceanography in La Jolla, California, and chairman of the NASA Advisory Council, admits that “there are all these fears out there” that Earth science will suffer under the new organization. For example,



Norwegian fjords as seen by NASA's Terra satellite. Researchers fear that streamlining at the agency will give less weight to Earth science projects.

NASA Earth science may no longer have an obvious advocate in interagency discussions on climate research, he says.

Lori Garver, a former head of NASA's policy office and now vice-president at DFI International, a Washington consulting firm, agrees that merging Earth and space science makes the former's budget “easier to raid” without having to consult Congress. But NASA's budget for Earth science was in decline even before the reorganization, and O'Keefe adamantly denied last week that it would suffer as a result of the change.

Strong advocate

Scientists inside and outside the agency also decried the loss of Weiler from the top science spot at NASA. Known as a tough decision-maker and a strong, articulate advocate for science, he will move to head the agency's Goddard Space Flight Center.

Although Weiler admits the switch was not his idea, he discounts speculation in Washington that he is being moved after clashing with O'Keefe over the administrator's decision not to have astronauts service the Hubble Space Telescope (see *Nature* 428, 353; 2004). He says that O'Keefe wants him to manage the proposed Hubble robotic servicing mission, one

of the agency's most challenging projects. He will take over a respected institution with more scientists than any other NASA centre, which has also just been given responsibility for the revitalized lunar science programme.

The new head of the merged Earth and space science directorate is Al Diaz, the current director of Goddard and a 40-year NASA veteran with senior management experience in Earth and space science. Diaz has strong ties with Senator Barbara Mikulski (Democrat, Maryland), who exerts a powerful influence on NASA's budget. But he is not a PhD scientist. Although Diaz has a solid managerial record, Berrien Moore, an Earth scientist at the University of New Hampshire and a frequent member of

NASA advisory groups, asserts that “the head of NASA science needs to be a distinguished scientist”. Nor has Diaz shown Weiler's flamboyance, or his flair for public relations. “He's not timid,” says a former colleague, “but he's not as ‘in-your-face’ as Ed is.”

However Earth and space science fare, the main purpose of the reorganization is to make the Moon–Mars programme a primary goal. Some critics question why NASA should make this change now, considering that Bush's plan got a lukewarm response in Congress and may not survive November's presidential elections. With NASA struggling to return the shuttle to flight, finish the International Space Station, and manage a full plate of science projects, there is “serious concern about confusion” if NASA's workforce now has to reorganize as well, claims one veteran agency observer.

But others say that a more streamlined agency makes sense. “Frankly, I like a lot of these changes,” says Garver, adding that they should improve communication within the agency and help clarify its missions to outsiders. “The human exploration programme will be much better off,” says Kennel, “and that's worth a lot to the agency, because half of its enterprise is human exploration.” ■

USGS/NASA

Kerry promises spectrum sale to fund science

Geoff Brumfiel, Washington

Democrat John Kerry, the US presidential candidate, has unveiled a strategy for investing billions of extra dollars in science and technology, with an emphasis on research in areas other than biomedicine.

The plan, released on 24 June, would raise US\$30 billion over an unspecified period of time through the auction of radio spectrum that is currently used for television broadcasts. Roughly two-thirds of the money from the sale would go towards research and development, and the rest would support incentives for industrial innovation.

But critics quickly panned the proposed source of funding as a financial gimmick. "Spectrum auctions are the last refuge of budget scoundrels," says Robert Walker, a lobbyist and former Republican chairman of the House of Representatives' Committee on Science.

Kerry announced the plan in a speech at San José State University in California. "America must lead, not follow, other countries in the great discoveries that bring greater prosperity," he said.

The proposal came at the end of a week of science-oriented campaigning for the Demo-

cratic candidate. At its start, Kerry said that, if elected, he would lift President Bush's ban on government-funded embryonic stem-cell research. He was then backed by 48 Nobel science laureates: "John Kerry will restore science to its appropriate place in government," they wrote in an open letter of endorsement.

The Kerry plan will auction radio spectrum freed for use by the conversion from analogue to digital television, says Thomas Kalil, a Kerry adviser and former assistant in President Bill Clinton's economic-policy office. The sale would produce more than \$30 billion; roughly \$22 billion would be assigned to research in areas such as nanotechnology and clean energy, and to science spending at the Department of Energy, the National Science Foundation and the Pentagon.

President Bush was advised to increase physical-sciences funding by his own scientific panel, but "has not followed through", says Michael Lubell, head of the American Physical Society's public affairs division in Washington.

Kerry's opponents say that his plan would be tough to implement and unlikely to produce dollars quickly enough. Direction of the funding would require an act of Congress, they note. And Walker points out that digital broadcasting of television is not due to be complete until the end of 2006, so no money would be available until then. Attempts to speed up the process would lead to a "huge argument with the broadcasters", Walker says.



Hand outs: the source of funding for John Kerry's science plan has been ridiculed by his critics.

Increasing virulence of bird flu threatens mammals

Helen Pilcher, London

An avian flu virus is mutating and becoming more dangerous to mammals, say researchers. The discovery reinforces fears that a pandemic could yet occur in humans.

Bird flu hit the headlines in 1997 when a strain called H5N1 jumped from chickens to people, killing six people in Hong Kong. Within three days, the country's entire chicken population was slaughtered and the outbreak was controlled.

Since then, new strains of virus have emerged, killing 14 more people. As yet, no strain has been able to jump routinely from person to person. But if a more virulent strain evolves, it could trigger widespread outbreaks, potentially affecting millions of people.

Now, genetic and animal studies show the virus poses an increasing threat to mammals. Immediate action is needed to stem the virus's transmission, says Hualan Chen from Harbin Veterinary Research Institute, China, who was involved in the research.

Chen and colleagues studied 21 H5N1

flu virus samples taken from apparently healthy ducks, which act as a natural reservoir for the disease, in southern China between 1999 and 2002. The researchers inoculated groups of chickens, mice and ducks with virus samples from different years. Their results are published this week (*H. Chen et al. Proc. Natl Acad. Sci.* doi:10.1073/pnas.0403212101; 2004).

As expected, the ducks were immune to the virus's effects and the chickens fell sick. However, the mice also became ill, losing weight and the use of their limbs. Crucially, the severity of their illness was linked with the year from which the virus sample was taken. Viruses isolated in 2001 and 2002 made the animals more ill than those isolated earlier on.

The findings hint that some time around 2001 the virus became adept at infecting mammals. Genetic analysis of the same samples reveals that the virus's DNA changed over that time, suggesting that accumulated mutations may have

contributed to the increased virulence.

Researchers are concerned that a virus that has acquired the ability to infect mice could also infect humans. "The disease could resurge at any time," warns virologist Marion Koopmans of the National Institute of Public Health and the Environment in Bilthoven, the Netherlands.

The findings highlight the need for improved surveillance to ensure that any future outbreaks are curtailed, she says. Although domestic poultry are easily culled, wild animals are more difficult to contain. "It is impossible to eradicate the natural reservoir," says Koopmans, "so we need to learn to live with it."

But birds may not be the only villains in this story. Chen believes that pigs may also play a part. In Asia, chickens and pigs are often kept in close proximity, so the virus may have shuffled back and forth between the two species, picking up mutations and becoming better at infecting mammalian hosts.



Field work: African subsistence farmers could benefit from more emphasis on agricultural research.

Research cloning gets green light from Japanese ethicists

David Cyranoski, Tokyo

Japan looks set to end its bar on therapeutic cloning, giving it the chance to compete with Britain, China and similar countries in fast-moving areas of biology such as stem-cell research.

On 23 June, the bioethics committee of the Council for Science and Technology Policy, Japan's main scientific decision-making body, voted to lift a three-year moratorium on therapeutic cloning by ten votes to five. The main council is expected to endorse the decision shortly. But rules to govern the procedure, which the committee will have to clarify before research can go ahead, will probably take a year or more to establish.

Therapeutic cloning involves creating an embryo by transferring the nucleus from a patient's cell to an egg cell stripped of its own nucleus. This clone yields stem cells that can develop into any of the body's tissues and may have the potential to repair the patient's organs. It promises much, but opponents say the research is advancing too quickly given ethical questions over its use of human eggs.

Shin-ichi Nishikawa, a committee member and stem-cell biologist at RIKEN's Center for Developmental Biology in Kobe, voted for the change, but worries that no plan has been agreed to allow research to start. "Considering the slow speed of decision-making in Japan, it is difficult to forecast when actual experiments with nuclear-transfer stem cells will be allowed," he says.

However, fellow committee member Motoya Katsuki, director of the National Institute for Basic Biology in Okazaki, says that the scientific and ethical problems that caused the moratorium to be put in place have not been solved. Katsuki, an embryologist who works with mice as models, opposed the lifting of the moratorium and says that abnormalities in cloned animals are still not properly understood. This could lead to wasteful experiments on human eggs, he adds. "From an ethical perspective we should avoid experimentation on human specimens as much as possible."

Controversy over the vote could erode vital popular support. Those against lifting the ban say there was not sufficient time for discussion, and some media reports have described the decision as a "forced consensus".

Over the next two weeks, the bioethics committee is expected to work out how to implement its decision.

UN urged to use science in fight against food shortfall

Helen Pearson, New York

Africa should pursue agricultural research more avidly in order to feed its growing population, scientific academies have told the United Nations. But the academies' recommendations are some way from filling empty bellies, development experts say.

"We need to mobilize the best scientists the world has to offer," said Kofi Annan, secretary-general of the United Nations, receiving the report on 25 June. *Realizing the Promise and Potential of African Agriculture*, was produced by the InterAcademy Council, an alliance of 90 scientific academies. It outlines ways in which scientists, farmers, industry and governments can work together to feed the 200 million malnourished people in Africa.

The academies propose that agricultural research should focus on the four farming systems that they think have the most potential: mixing maize and animals; combining cereals and roots; irrigated systems; and tree crops.

They urge African nations to support partnerships between publicly funded research groups and private companies to ensure, for example, that new strains of crops win market access. This is seen as essential in providing farmers with secure income and helping them to escape subsistence farming. The academies also call for the foundation of internationally funded 'centres of excellence' in Africa and for a doubling in the agricultural research spending of both African governments and donor agencies by 2015.

Agricultural research experts broadly welcomed the attempt to shift public attention away from emergency famine responses towards securing Africa's food supply.

But some question how the recommendations will be translated into action and who is going to cough up the cash. "That," says

Yvette Stevens, a UN adviser on development in Africa, "is where these things drop dead."

Others point out problems with the specifics of the report. For example, scientists at the proposed centres might be more interested in publishing papers than addressing local problems, says Richard Jefferson, head of the Center for the Application of Molecular Biology to International Agriculture in Canberra, Australia.

Deeper problems in trade policy and African infrastructure need to be addressed before science can benefit farmers, say many development specialists. Agricultural subsidies in rich countries, for a start, make it almost impossible for African crops to compete on price, at home or abroad. "If there is one intervention that would transform Africa's potential," says Jefferson, "it would be the complete abandonment of agricultural subsidies."

Even when agricultural research succeeds in boosting crop yields, lack of roads and other basic infrastructure can prevent the resultant harvest getting to market. These issues need to be addressed alongside the scientific ones, says Peter Hartmann, director of the International Institute of Tropical Agriculture in Ibadan, Nigeria.

But Calestous Juma, a development specialist at Harvard University, welcomes the report and believes the next challenge is to convince the heads of African governments: "Without high-level political commitment there is not pressure to realize these goals."

The Forum for Agricultural Research in Africa is drawing up regional action plans based on the report, says Rudy Rabbinge, a development specialist at Wageningen University in the Netherlands and co-chair of the report panel. Annan will personally convey the findings to heads of state at the meeting of the African Union in Addis Ababa, which started this week.

Weapons plan gives United Nations key role

Geoff Brumfiel, Washington

Arms-control specialists have proposed a sweeping plan to put the brakes on the international spread of nuclear weapons.

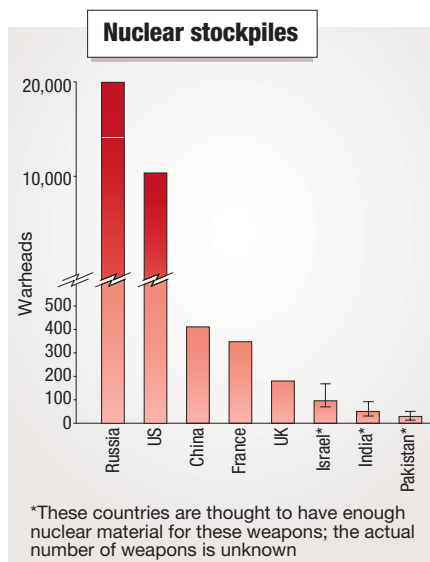
The Carnegie Endowment for International Peace, a Washington-based think-tank, unveiled its plan to government officials and other specialists at its non-proliferation conference in Washington on 21 June.

The proposal, which is put forward as existing non-proliferation arrangements come under pressure in Iran, North Korea and elsewhere, relies on the United Nations taking a more proactive role in preventing the spread of nuclear weapons and materials.

Current approaches to non-proliferation are not working, says Joseph Cirincione, director of the Carnegie Non-Proliferation Project and a co-author of the plan. "We hope this report will build an expert and political consensus around a new strategy," he adds. But critics say that the draft report is a hotchpotch of sometimes conflicting ideas that cannot easily be combined.

"In my view, this could be the beginning of a much needed discussion on non-proliferation and security," Mohamed ElBaradei, director-general of the International Atomic Energy Agency, told the meeting.

Opinion in the United States is sharply divided between those who support a system of arms control based on international treaties, and those who think that the United States should use its military and political power to stop the spread of nuclear weapons.



The Carnegie plan outlines five basic duties for nations. They must work to ensure that there are no new nuclear-weapons states, secure existing stockpiles of nuclear materials, stop illegal transfers of nuclear-related technology, commit to resolve conflicts in their regions, and diminish the overall number of nuclear weapons. Current estimates put the number of nuclear weapons worldwide at more than 30,000, with the majority in Russia and the United States (see graph).

The plan calls for "universal compliance" from all nations, whether they are party to existing nuclear-weapons treaties or not. Any

nation participating in the illegal transfer of technology, or actively pursuing its own programme to build a bomb, would be subject to sanctions or military intervention from the United Nations Security Council.

"There is a lot of good stuff in the report," says Michael Levi, an arms-control expert at the Brookings Institution in Washington. But he adds that some of the ideas will be tough to sell. A proposal to halt the production of weapons-grade uranium worldwide, for example, is unlikely to go down well in Europe, where nuclear power plants produce such material as a by-product.

Others object to sections of the report calling for the United States and Russia to stop nuclear testing and to halt research on new nuclear weapons. "Developing new types of nuclear weapon may be essential to deterrence," says Baker Spring, a defence analyst at the Heritage Foundation in Washington.

"A lot of ideas that are not very comfortable alongside each other have been stuffed into this report," says Paul Robinson, director of Sandia National Laboratories in Albuquerque, New Mexico, the main US nuclear-weapons engineering lab. Among them, he says, is a proposal to allow Pakistan, India and Israel to join the nuclear non-proliferation treaty. This would reward them for defying the 1970 treaty, Robinson contends.

Cirincione says that he and his co-authors welcome different views on the subject, and will release a "significantly different" final plan in December.

Santa Cruz brush-off leaves artists in a different class

Jonathan Knight, San Francisco

An unusual and highly regarded programme to train scientific illustrators at the University of California (UC) has suffered a setback under state budget cuts. Over the summer it will move from the university's campus in Santa Cruz to its less prestigious extension programme nearby.

The switch illustrates how the less well-known aspects of the scientific enterprise suffer first in lean times. "It's an amazing programme. It has produced brilliant artists," says Malcolm Margolin, a publisher of natural-history books in Berkeley, California. "But illustration is a neglected art in the age of photography."

Science illustrators occupy a niche in science communication that photography cannot fill. For instance, in drawing wildlife they can highlight several features of an animal, such as the details of a feather and the shape of an eye, that a photographer would have difficulty capturing in a single

image. Illustrators can also render extinct creatures — the four-winged dinosaur on the cover of *Nature* on 23 January 2003, for example, was drawn by a graduate of the Santa Cruz programme.

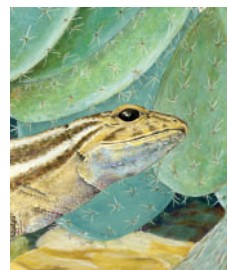
Illustrators can also draw conceptually difficult processes that are too small to see directly, such as protein synthesis or chemical catalysis. They can produce vibrant landscapes from the murky footage returned by submarines in deep ocean trenches or the transmissions of a martian lander. Their drawings illustrate textbooks, museum exhibits, scientific papers and popular magazines. But the Santa Cruz programme is one of only a handful to train such people.

The programme's move was forced by state budget cuts, says David Kliger, dean of physical and biological sciences at UC

Santa Cruz. It was more expensive than other programmes because it retained three faculty members to teach only ten students. "It was a luxury we couldn't afford right now," he says.

Although Kliger says he does not expect the move to adversely affect the programme, former students worry that reduced access to campus libraries and labs may be a problem. "It was important to be on campus," says 2004 programme graduate Amadeo Bachar, now an intern at *National Geographic* in Washington. "At least half of us were working with research professors."

The move may already have cost the programme some applicants. Only 20 applications had arrived by the 15 May deadline, less than half the usual number, says Andrea Michels, manager of the science-communication department.



Top drawer? Scientific illustration can be used to emphasize details.

Junior biologists score partial victory over lab conditions

UNIV. KONSTANZ

Alison Abbott, Munich

For a group of junior biology researchers in Germany, proof has arrived that there is indeed strength in numbers.

In a final bid to get their voices heard, the group last spring submitted a joint complaint about their alleged mistreatment in a lab at the University of Konstanz. An internal investigation by the university has now supported some of their grievances.

The committee charged with investigating the researchers' claims finished its report on 24 May. It found that the lab's head, evolutionary biologist Axel Meyer, worked his team too hard, and that "occasionally insufficient" support "led to frustration or overtaxing" of the lab's junior members.

The committee's report stresses that the junior researchers made no accusations of data manipulation. But it states that some of the complaints hinted at scientific misconduct. The report, which was leaked to media outlets by one of the complainants, adds that Meyer's management style "could be detrimental" to the scientific activity of his group.

Observers of the intensely hierarchical German university system say that the case is probably the first time that junior scientists have successfully joined together to make a complaint against a professor.

Meyer, who joined the university in 1997, denies all charges of inappropriate behaviour. He says that incidents in the laboratory have been "blown out of proportion" in the report.

Meyer's busy and productive lab usually hosts about a dozen PhD students or post-docs from around the world. Some who passed through the lab say that they found the atmosphere scientifically stimulating; but others were unhappy and many of them left before the end of their contracts.

Finding little success in bringing individual complaints to the university, 16 researchers — all but six of whom have now left Konstanz — grouped together and presented a joint complaint in May 2003. In September the university asked its standing committee on responsibility in science to look into the case. The panel is chaired by law professor Dieter Lorenz and investigates scientific practice under guidelines set by the DFG, Germany's main research funding body.

The committee, which called on the advice of outside experts in Meyer's field, considered only a sample of the complaints filed. It concluded that Meyer had been insufficiently concerned about the welfare of



The University of Konstanz had complaints from 16 researchers.

the young scientists in his lab. In one extreme case, the report says, the nervous breakdown of a foreign PhD student "could be seen as an indicator of the general negative situation which developed from the optimization of scientific output".

A number of specific accusations were selected by the committee for investigation, and it decided that misconduct had occurred in four cases. These included advertising positions for which independence of research was promised but not delivered, and for which there was no specific budget. In a statement made through his lawyer, Meyer said that his assistants had "comparable freedom with non-tenure track US assistant professors".

The committee supported the students' allegations that Meyer also demanded honorary authorship on papers and that in one case he used "almost word for word" a grant proposal designed and written by a foreign scientist to the European Commission's Marie Curie postdoc programme for an application to another granting body. The young scientist involved claims that she arrived at the lab to find another student working on the project, and that she then had to work on something different. Meyer counters that he was deeply involved in developing the idea with her from the outset.

The panel found no evidence for bad practice in five other allegations it investigated. The young scientists say that they are pleased the committee has accepted some of their complaints, but that they still stand by the rejected claims.

Supporters of Meyer have rallied to his defence. "It was not an easy time, but it was the most interesting time in my life," Miguel Vences, now at the Zoological Museum in Amsterdam, says of his time in Meyer's lab. "Before then, I was not used to thinking so big."

The university declined to comment on how the case will now be taken forward. Its rector, Gerhart von Graevenitz, is expected to decide shortly what action should be taken. ■

Academics seek to cast peer review as a public service

Declan Butler, Paris

Scientists have been rapped on the knuckles by a panel of academics who spent more than a year assessing public awareness of peer review.

The panel's report, published on 24 June, says that researchers and their institutions should explain the importance of peer review to the public to help them to weigh up scientific claims. Every time scientific results are presented in public, it should be made clear which parts of the claims have been peer reviewed and which have not, the academics say.

"A culture of explaining and asking about peer review all along the line — from radio phone-ins to ministerial briefings — will put a lot more pressure on people to explain exactly what the status of the work is," says Tracey Brown, director of the London-based watchdog Sense About Science, which produced the report.

The Sense about Science panel included Colin Blakemore, head of the UK Medical Research Council; Peter Lachmann, an immunologist at the University of Cambridge; and John Maddox, former editor of *Nature*.

Their report lists a litany of scares that had their origin in publicized but unpublished claims, including the health risks of radiation from mobile phones, the fear that the MMR (measles, mumps and rubella) vaccine could cause autism, and that acrylamide in fried foods could cause cancer. Such scares could be killed at birth if journalists and politicians paid more attention to the publication status of the original claims, says Brown.

That's an ambitious goal. A poll commissioned this year by the Science Media Centre and *Nature* — conducted by the London-based market-research company MORI — showed that almost 75% of the public don't know what peer review is. Sense About Science will work with research and educational bodies to encourage teaching about peer review in schools and universities, says Brown.

But scientists themselves are the report's main target. Researchers often whine about the hassles of peer review, complaining about the practicalities and potential flaws of the system, says Brown. Instead, she says, scientists should use peer review to explain to the public the validity of scientific claims. "Here is something in scientists' armoury to help the public understand science better, but they never think about peer review apart from moaning about it," says Brown. ■



Behind bars: Bulgarian nurses in court during their trial in Libya in January.

Montagnier speaks out for medics under death sentence in Libya

Washington A group of scientists from around the world is making a last-ditch plea to the Libyan government to spare the lives of five Bulgarian nurses and a Palestinian doctor who have been sentenced to death.

In May, the healthcare workers were convicted of deliberately infecting 400 children in a Libyan hospital with HIV, and sentenced to death by firing squad.

Luc Montagnier, one of the co-discoverers of HIV, investigated the case with Vittorio Colizzi of Tor Vergata University in Rome. They concluded that the virus was more likely to have been spread through unsterile hospital practices than by deliberate infection.

On 23 June, Montagnier and a group of scientists sent a letter to Libyan leader Colonel Muammar Gaddafi asking him to commute the medics' death sentences. The letter, organized by Physicians for Human Rights in Boston, Massachusetts, included signatures from the other co-discoverer of HIV — Robert Gallo of the Institute of Virology in Baltimore — and 28 other scientists from Egypt, Iran, France, the West Bank and Gaza, and four other countries.

Libya has so far not responded to pleas for leniency from nations such as the United States, says Montagnier, but he hopes they will listen to the scientists' protest. "This is a very important issue of human rights," Montagnier says.

Lawyers are expected to appeal against the sentences in mid-July.

Don't mention 'therapeutic cloning', society says

New York Scientists working on human cloning are being urged to change their vocabulary. The International Society for Stem Cell Research is asking its members to use the phrase 'nuclear transfer' instead of 'therapeutic cloning' in future papers and in communications with the public and press,

after discussions at its annual meeting in Boston in June.

Experts say that 'nuclear transfer' is a more accurate term for the process, which involves inserting a nucleus from one cell into an egg stripped of its own genetic material.

The society says the switch is also aimed at avoiding widespread confusion among the public and legislators between 'therapeutic cloning', in which the resulting embryo is used to make stem cells for medical research, and 'reproductive cloning' to make babies. "You can end up banning by accident things you didn't mean to ban," says stem-cell researcher Lawrence Goldstein of the University of California, San Diego.

Meeting instrumental to saving science history

New York Do you have a load of antique scientific instruments gathering dust in a university storeroom? If so, Francis Manasek, who chaired a meeting on scientific collections last week at Dartmouth College, New Hampshire, would like to hear from you.

Instruments that figure prominently in major discoveries — such as the first clock used to measure longitude at sea — usually make their way into museums. But Manasek, a research scholar in physics and astronomy at Dartmouth, is keen to preserve less famous objects, too. "They aren't just pretty items," he says. Even simple slide rules or microscopes can help to illuminate, among other things, the evolution of scientific teaching, he says.

Manasek hopes that the meeting — at which some 80 participants reviewed practical matters such as laws for displaying artefacts and handling mercury — will help to spread the preservation movement.

US and Europe agree a position on global navigation

London The United States and the European Union (EU) have signed a deal to cooperate over satellite navigation, settling a long-running dispute over the EU's planned Galileo satellites.

Europe has long been keen to have its own satellite navigation system, independent of the US Global Positioning System (GPS). But the system it initially proposed for its Galileo satellites had the potential to interfere with a signal used by the US military, leading the United States to protest.

Both parties have now agreed to use the same system for new navigation satellites. This has slightly less precision than the one originally favoured by Europe. But by both systems operating together, experts say they will gain greater accuracy and reliability.

Medical highlights get a dusting off online

London Papers from some of the world's oldest medical journals, dating back over the past 125 years, are to be made freely available online.

Work from about 15 journals will be put on PubMed Central, a digital archive run by the US National Library of Medicine, thanks to £1.25 million (US\$2.3 million) granted by the UK government and the Wellcome Trust — a London-based medical charity.

Famous papers encompassed by the project include the 1939 report in *Annals of Surgery* of the successful repair of a faulty heart artery in a newborn baby, and research from the 1880s that led to the first solutions used to keep cells alive in culture.

The papers should start to appear online by early 2005.

High-protein study hints at fertility problems

Berlin Women who follow the Atkins diet may be reducing their chances of becoming pregnant, researchers reported at the European Society for Human Reproduction and Embryology meeting in Berlin on 28 June.

"The data indicate that a high-protein diet is not advisable while trying to conceive," says David Gardner of the Colorado Center for Reproductive Medicine in Englewood, Colorado, who led the study.

The popular Atkins slimming diet restricts carbohydrates and includes more protein than is generally recommended by dietitians.

Gardner's team found that mice fed a diet packed with 25% protein before conception were less likely to produce viable offspring. About 15% of their embryos were miscarried,

compared with just 1% of embryos in mice fed a standard diet of 14% protein.

The researchers think that the high-protein diet may alter genetic imprinting — the process by which genes are stamped as being from the mother or the father, which affects their expression in the embryo.



From rags to research

As a child, geneticist Mario Capecchi had to fight for survival after his mother was imprisoned by the Nazis. He tells his story to Carina Dennis, hoping to inspire others from a disadvantaged background.

Hunger focuses the mind. And hunger is what Mario Capecchi remembers most from his early childhood, which was spent on the streets of war-torn Italy. Today a contender for a Nobel prize, the geneticist attributes the single-mindedness and fearlessness that has underpinned his scientific success to this formative struggle for survival.

Capecchi, who works at the University of Utah in Salt Lake City, is fêted for his contribution to methods for selectively knocking out genes in mice — widely used to investigate gene function and to create animal models to study human disease. Yet while he was conducting the experiments that made his name, none of Capecchi's peers knew of his harrowing childhood experiences.

When he spoke out for the first time, while accepting a Canadian prize for medical science in 1993, Capecchi shocked even his wife, who had heard only snippets of the tale. "Subsequently, I wrote about the story when I received the Kyoto prize in 1996," he says. "To make sure that there were no more surprises I showed her a draft of the manuscript."

Capecchi's life story is humbling. Born in Verona in 1937, he was the product of a brief relationship between Lucy Ramberg, a poet, and Luciano Capecchi, an Italian air-force pilot. Nearly four years later, his mother — who was one of the Bohemians, a group of artists outspoken in their opposition to the Nazis — was taken away by the Gestapo and interned in the Dachau concentration camp.

Before her arrest, Capecchi's mother sold all her belongings and used the money to arrange for a peasant family to care for him. But for reasons Capecchi has never understood, the cash soon ran out. At just four-and-a-half years of age, he was turned out onto the streets. Although he is happy to talk in general terms about his early life, Capecchi refuses to be drawn into the grim specifics. "The reality was brutal and there is little to be gained from recounting the details," he says.

Slow start

After moving to the United States when the war ended, Capecchi started at elementary school aged nine without knowing a word of English or how to read, write or do mathematics. His teachers said he would never go to college. By talking about his troubled beginnings, Capecchi hopes to encourage lab heads not to dismiss the potential of people from disadvantaged



Mario Capecchi hopes that his life story will help disadvantaged people to fulfil their potential.

backgrounds. "My message is that anyone in any circumstances can make it," he says.

Capecchi has proved his teachers wrong. So, too, the dismissive reviewers of a grant proposal he submitted to the US National Institutes of Health in 1980, detailing his strategy for disabling genes in mammalian cells. It was too speculative, they ruled.

Undeterred, Capecchi continued the research, diverting funds from other projects. His efforts were rewarded. Capecchi's was one of two groups — the other led by Oliver Smithies, then at the University of Wisconsin — that a few years later showed that a dysfunctional copy of a mammalian gene will, on rare occasions, line up with and replace the normal version, in a process called 'homologous recombination'^{1,2}.

Knockout mice are created by performing this trick in mouse embryonic stem (ES) cells. The transformed cells are injected into embryos, and breeding from the resulting mice produces animals that carry the disabled target gene in all of their cells.

Martin Evans, then at the University of Cambridge, UK, had by the mid-1980s iso-

lated mouse ES cells³ and showed that they could be incorporated into a developing embryo to create a 'chimaeric' animal⁴. Capecchi demonstrated that homologous recombination also happens in ES cells⁵, and developed efficient methods to select the tiny proportion of cells in a culture to have under-

gone this transformation⁶. The stage was set for the debut of targeted knockout mice, first described in 1989 and 1990 by groups including those led by Capecchi and Smithies⁷⁻¹¹.

Capecchi ascribes his perseverance with this work, despite the initial scepticism of his peers, to his formative experiences on the streets. His childhood fixation on finding food fostered an ability to concentrate on the problem at hand, regardless of distractions. "It's both a strength and a weakness," says Capecchi. Although he can focus intently on a complex research question, he isn't a multi-tasker. "He thinks all these big thoughts but he can't remember where his glasses are," observes one of his lab staff.

When he was left to fend for himself, Capecchi survived by begging and stealing. He

"I think it can be an advantage being naive, because you can come up with ideas that haven't been thought of before."



Shattered streets: Italian children cluster round as Allied soldiers hand out sweets in 1944.

often travelled with other children, operating in small groups. One would create a diversion while the others stole food and other supplies. "But I couldn't stay in one location for very long, because fairly soon everyone would recognize me and what I was up to," says Capecchi. Occasionally he was picked up and put in a hospital or orphanage — mostly brutal institutions from which he soon escaped.

Seven-year itch

His nomadic beginnings are reflected in his science. Capecchi likes to change fields roughly every seven years. He worked on bacterial viruses before moving into the mammalian genetics for which he is renowned. Today, he is interested in the developmental genetics of the nervous system and behaviour.

These transitions required Capecchi to attend conferences as a relative unknown, swot up on new literature, and face other scientists wary of newcomers encroaching on their turf. "They often don't welcome you with open arms," he says.

Given the real horrors that Capecchi has endured, he has never been scared to seem ignorant. "I think it can actually be an advan-

tage being naive, because you can come up with ideas that haven't been thought of before," says Capecchi. "And the broader your experience is, the less fearful you are to try to do different things."

Capecchi credits his love of pursuing big questions to James Watson, co-discoverer of the DNA double helix, who was his PhD supervisor at Harvard University. "He is the most important person in my scientific career," says Capecchi.

But Capecchi's relationship with his famously outspoken mentor wasn't without friction. He recalls the time they disagreed over the results of an experiment on the rates at which the subunits of ribosomes, the cellular bodies that make proteins, come together and separate. Capecchi was not convinced by the data and suspected that an unknown protein might be involved. He wanted to repeat the work; Watson thought the case was closed. When the argument escalated, Capecchi threw glass plates bearing key data into the bin, ensuring that he would not have to publish the results.

Colleagues ducked for cover as Watson exploded in fury and showered Capecchi with expletives. "I came that close to being

thrown out of the lab," Capecchi says. But he was right — a few years later a ribosomal dissociation factor was discovered¹².

Meeting the quietly spoken Capecchi today, it is difficult to imagine him being confrontational. He shuns large conferences, and prefers to keep his research group to a manageable size of about ten postdocs and five graduate students. "I don't do well in large groups," he admits.

Joint effort

Within this cohesive unit, he encourages a spirit of social interaction and mutual assistance inspired by his adolescent years in a Quaker community. After the war, Capecchi's mother tracked him down on his ninth birthday — she found him in hospital, naked and half-starved. He didn't recognize her. "She had aged a lifetime," he says.

Three days later, the pair boarded a ship bound for the United States to live with Capecchi's uncle, who with his wife had founded a Quaker community north of Philadelphia. Capecchi's mother never recovered from her wartime experiences. "She lived in her imagination for the rest of her life," he says. So it was left to his uncle and aunt to bring up the young boy.

Capecchi's experiences have left him fiercely self-sufficient, and keenly aware of the importance of supportive and inspiring mentors. "He is very open to new ideas and will almost never say no if something broadens the scope of the question," says Matthew Hockin, a postdoc in Capecchi's lab.

When choosing new members of his team, Capecchi applies his philosophy of looking beyond people who have the obvious scientific backgrounds. He puts much more emphasis on applicants' drive and passion, rather than their curriculum vitae. "What I'm mainly looking for is curiosity and excitement," he says.

"Your legacy is who you have trained. I think of them as an extended family who I hope will continue an interesting line of questioning that I helped foster," Capecchi adds. Given his own experiences, he is much more interested in where his protégés end up, rather than where they started. ■

Carina Dennis is Nature's Australasian correspondent.

1. Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. *Nature* **317**, 230–234 (1985).
2. Thomas, K. R., Folger, K. R. & Capecchi, M. R. *Cell* **44**, 419–428 (1986).
3. Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154–156 (1981).
4. Bradley, A., Evans, M. J., Kaufman, M. H. & Robertson, E. J. *Nature* **309**, 255–256 (1984).
5. Thomas, K. R. & Capecchi, M. R. *Cell* **51**, 503–512 (1987).
6. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348–352 (1988).
7. Schwartzberg, P. L., Goff, S. P. & Robertson, E. J. *Science* **246**, 799–803 (1989).
8. Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. & Jaenisch, R. *Nature* **342**, 435–438 (1989).
9. Koller, B. H. & Smithies, O. *Proc. Natl Acad. Sci. USA* **86**, 8932–8935 (1989).
10. Thomas, K. R. & Capecchi, M. R. *Nature* **346**, 847–850 (1990).
11. McMahon, A. P. & Bradley, A. *Cell* **62**, 1073–1085 (1990).
12. Subramanian, A. R. & Davis, B. D. *Nature* **228**, 1273–1275 (1970).

Breezing into town

The world's highest urban wind farm could be a flagship project for renewable energy, if New York City planners get their way. Are city dwellers ready for wind power? Jonathan Knight investigates.

Everyone remembers where they were on 11 September 2001. Building engineer Sinisa Stankovic was at the Rutherford Appleton lab near Oxford, UK, for a day-long workshop on renewable energy. He and his colleagues had just presented results from a two-year study on how to integrate wind turbines into city buildings, when news of the terrorist attacks came in.

Now, in a strange twist for Stankovic, the planned redevelopment of Ground Zero in lower Manhattan might give a boost to urban wind power. The site's signature structure, the Freedom Tower, is scheduled for a ceremonial groundbreaking on 4 July, American Independence Day, and is also set to host the world's tallest wind farm.

If built as planned, the tower would be the world's most visible symbol of renewable energy, with as many as 30 spinning wind turbines in its upper storeys, providing a fifth of the building's power. "The Ground Zero turbines will be high-profile, and that might spur a lot more interest in wind," says Stankovic, who is a founder of the London-based ecological design consultancy BDSP Partnership.

So is this the shape of things to come? Many wind-power enthusiasts say that rooftop generation will play a role in the future, as fossil-fuel supplies dwindle and customers seek greater independence from the creaking electrical power grid. Others doubt whether urban wind power will have much success, and some are even sceptical that the Freedom Tower turbines will be built.

Today, most wind generation occurs at wind farms in remote rural areas or offshore, where the wind blows on average three times stronger than in cities. These giant turbines, with rotating blades 20–60 metres long, can each generate up to 5 megawatts, enough to power a large office building.

The largest turbines in cities are also free-



standing units. Toronto boasts a 30-storey lakeside windmill that can power up to 250 houses when spinning. Unlike rural wind farms that are often opposed by residents on aesthetic grounds, Toronto's turbine has the enthusiastic support of the local community, who raised 50% of the installation costs.

Winds of change

Yet rooftop windmills are rarely seen, despite the fact that wind turbines are cheaper to install and run than solar panels. According to wind-power experts, building owners have concerns that turbines might be too noisy or heavy. Others consider the huge spinning blades unsightly or even dangerous.

Enthusiasts for urban wind power believe these problems can be overcome. "We think wind power fits better into a man-made environment than a pristine one," Stankovic says.

With limited space for rural wind farms, Europe has taken the lead in research into urban wind power. The European Union expects member nations to get some 12% of their energy from renewable sources by 2010, and Britain is seeking to derive at least 20% from renewables by 2020. Achieving these goals will require less centralized production, with consumers using solar panels or wind turbines to generate some of their own power.

In 1998, Stankovic helped to coordinate a project funded by the European Commission to explore ways of incorporating wind power into cities. Project WEB (Wind Energy for the Built Environment) concluded that buildings designed or retrofitted with wind turbines need to get at least 20% of their electricity from wind to justify the installation cost. To help achieve this, the shape of the building should be designed to

M. ARAD/P. WALKER/LMD/C



Gust stars: Toronto has a giant wind turbine (above) on its outskirts, but plans for New York's Freedom Tower (left) may see smaller, vertical-axis turbines (right) installed in the upper storeys.

maximize the efficiency of the machines.

Square buildings disrupt the smooth flow of air and create turbulence, which lowers turbine efficiency, the project concluded. So structures that incorporate turbines need curved surfaces or ducts to keep the wind flowing smoothly towards the rotor.

Gale forced

Project WEB participants built a two-storey design prototype to show that a turbine integrated into the space between two towers can be 25% more efficient than a free-standing device¹. The footprint of each tower forms a boomerang shape that accelerates air flow over the turbine and stops turbulence. "By concentrating the wind," says Stankovic, "buildings can enhance the efficiency of turbines."

The Freedom Tower architects, Skidmore, Owings & Merrill, claim that the open latticework structure of the tower's upper section will allow air to flow unimpeded over the wind turbines. They also claim that the building's shape twists to face the prevailing winds, although few details are given.

But critics note that large turbines in cities pose several problems, not least the danger to people and property should a blade break. Blades can weigh hundreds of kilogrammes, and tip speeds can exceed 100

kilometres per hour, so an escaping blade fragment could do considerable damage.

Urban-wind proponents say that such fears are unfounded. Modern turbines have safety features including speed control and blades that turn edge-on into the wind during storms. Perhaps the biggest challenge for building-mounted turbines is noise.

Any slight imbalance in the blades is amplified by centrifugal forces, causing the turbine to shake when rotating. "No rotating machinery is perfectly balanced," says Clint 'Jito' Coleman, president of Northern Power Systems in Waitsfield, Vermont, and a 30-year veteran of the wind-power business. "Any imbalance shows up once per revolution, and there is nothing you can do about it."

The noise produced can vary from a regular thump to a low rumble. Should the turbine speed match the harmonic resonance frequency of the surrounding structure, such as a supporting beam, then the building itself can also vibrate, amplifying the sound.

"They can turn your roof into a big loud-speaker," says Mike Bergey, who runs Bergey Windpower in Norman, Oklahoma, one of the oldest wind-power companies in the United States.

If the turbine is part of the original design, as in the Freedom Tower, it should be possible to dampen vibrations by constructing heavy support structures, Bergey says. But that gets expensive. When the architects began looking for a wind-power company to design their turbines earlier this year, both Bergey and Coleman decided not to bid, believing that the cost would be prohibitive. "I'm sceptical that there will ever be turbines up there," Bergey says.

A spokeswoman for Skidmore, Owings & Merrill declined to answer *Nature's* questions about the Freedom Tower turbines or to say whether an engineering firm had been chosen to build them. But one wind-power company that is bidding on the Freedom Tower project says that it has the solution to vibrations: vertical-axis turbines. Unlike traditional windmills with their propeller-like blades, vertical turbines have curved blades that are attached at both ends to a vertical shaft. They spin whatever the wind direction.

Gary Westerholm, president of McKenzie Bay International in Brighton, Michigan, says that vertical turbines produce less vibration at a given rotation speed because the blades do not stick out so far and so exert less pull.

To reduce the vibrations still further, Westerholm proposes installing 30 small, 100-kilowatt, turbines in the Freedom Tower, rather than fewer large ones. The turbines also automatically change speed if they approach resonance frequencies of any components.

Despite these advantages, vertical turbines have not been commercially successful. The

few firms pursuing them in the 1970s and 1980s ran into financial problems, says Robert Thresher, director of the National Wind Technology Center in Golden, Colorado. "It was basically bad luck," he says. Even today, there are few installations of vertical turbines.

Vertical challenge

Most low-rise city buildings don't have room for 30 turbines, whatever their size, and suffer from weak and unreliable winds. So a pilot project in Scotland is exploring ways to make small turbines more attractive to architects by combining them with other technologies that reduce a building's power needs.

Designed by Charles Rennie Mackintosh, Scotland's best known architect, Glasgow's 1895 Lighthouse Building is a national treasure. When it was refurbished in 1999, engineers at the Energy Systems Research Unit (ESRU) of the University of Strathclyde were consulted about adding wind and solar power.

The redesign added openings just below the roofline to channel air up through ducts, across turbine blades and out onto the roof. Ducts offer several advantages. By channelling the wind, they reduce air turbulence before the air meets the rotor. They also allow the turbines to be hidden away, and prevent broken blades from escaping and doing damage, says ESRU director Joe Clarke.

The Lighthouse Building's other efficiency features include windows that keep heat in during winter and out in the summer, motion sensors that ensure only occupied rooms are heated, and lights that dim if there is enough daylight available. Combined, these measures make the building 70% more efficient than the UK best-practice standards require. As a result, the turbines can produce up to a third of the building's annual power consumption².

Although no projects as ambitious as the Freedom Tower have yet been built, the number of small windmills is on the rise. In the Netherlands, dozens of them spin on rooftops in The Hague and Amsterdam. And from this May, the Scottish government is funding a trial of small turbines on the roofs of primary schools in Fife.

Tourists visiting the viewing platform of the Freedom Tower won't fail to notice its patriotic symbolism. The finished tower will rise to 1,776 feet (545 metres) — a reference to the year of American independence — and its capping spire is supposed to mirror the Statue of Liberty's upraised arm. But will visitors be surprised to see windmills among the skyscrapers? By 2015, perhaps not. ■

Jonathan Knight writes for *Nature* from San Francisco.

1. Campbell, N. et al. in *Proc. European Wind Energy Conf. & Exhibition (EWEA)*, Brussels, 2001; <http://www.bdsp.com/web>
2. Clarke, J. A. et al. in *Proc. 16th European Photovoltaic Solar Energy Conf.* (James & James, London, 2000); ftp://ftp.strath.ac.uk/Esru_public/documents/LH-PV.pdf

Change of mind

A brain haemorrhage turned an ex-convict into an obsessive artist. Jim Giles meets him and the scientists studying his case.

Tommy McHugh was sitting on the toilet when his life changed. He felt a sudden and extreme headache, but something far more serious was occurring: blood was leaking from an artery in his brain. McHugh was suffering from a cerebral haemorrhage — an event that hospitalized him, altered his personality and, he says confidently, was the best thing that ever happened to him.

Standing in the bright London sunshine three years later, McHugh describes his new life. If I wasn't there, he tells me, he would be manically painting, sculpting or writing poetry. It's a statement those who knew McHugh before his stroke would find hard to believe. A former builder and heroin addict who has had stints in jail for violent offences, McHugh had no former interest in art. Now he spends almost all his time compulsively creating. "My life is 100% better," he grins.

McHugh's story is extraordinary but not completely unique. A handful of similar cases have been identified by neurologists in the United States. There are even rumours that Tom Cruise may play the part of one such patient in a forthcoming film. While the number of subjects remains small, only tentative conclusions can be drawn about how these artistic urges occur. But studies of patients such as McHugh could shed light on how our brains create art. "People like Tommy could reveal the key to artistic creativity," says Mark Lythgoe, a neuroscientist at University College London (UCL).

Hidden from view

It is hard going for Lythgoe, though — in part thanks to limitations on the methods he can safely use to investigate McHugh's brain. Doctors stemmed McHugh's bleeding by clamping a metal clip on the damaged artery. So he cannot be scanned using magnetic resonance imaging, as the magnetic fields could move the clip and cause further brain damage. Instead McHugh has had only computerized tomography scans, taken before and after the operation. These X-ray scans aren't ideal for picking up subtle signs of brain damage — in McHugh, they show nothing unusual at all.

But McHugh's art reveals some aspects of his altered thought processes. He often draws images of faces within faces, which he describes as representing two parallel lines of thinking that seem to run through his brain simultaneously.

In cognitive tests conducted by UCL psy-



Creative urge: Tommy McHugh works on another sculpture in his art-packed living room in Liverpool.

chologist Michelle de Haan, McHugh demonstrates a mix of abilities and deficiencies. He scores in the normal range for IQ, but he has difficulty switching trains of thought. When asked to list types of furniture, for example, he will rattle off chairs and tables like the rest of us. But when asked to switch to listing animals, he continues to talk about lamps and sofas. It's a trait that becomes apparent when talking to McHugh: sentences tumble out relentlessly unless he is interrupted.

Similar deficits are seen in patients with damage to their frontal lobes, a region of the brain associated with high-level functions such as planning. This suggests McHugh may likewise be injured here.

Artistic inspiration

That would make McHugh similar to patients studied by Bruce Miller of the University of California, San Francisco — one of the few neurologists to take an interest in the link between art and brain damage. Miller's patients all have semantic dementia, a condition in which sufferers lose the ability to communicate through speech, caused by damage to the frontal and temporal lobes of the brain. About 20 such patients seen by Miller since 1990 also have a new and sometimes compulsive artistic interest or skill. In this set of patients, the damage is predominantly to the left side of the brain (B. L. Miller

et al. Br. J. Psychiatry **176**, 458–463; 2000).

Areas on the right side of the brain are usually involved in artistic functions such as visual processing, face recognition and spatial awareness. But these are moderated by areas on the left-hand side. Perhaps, speculates Miller, damage to the left lessens this inhibition and causes patients to become more interested in art.

McHugh's case is slightly different, as he has no speech impairment. But to understand exactly what is going on in McHugh's brain, Lythgoe will need to find others with the same symptoms. "I have a feeling there are other people like Tommy out there," Lythgoe says. "But unless the neurologists who are treating them have an interest in art they often don't get followed up."

In the meantime, Lythgoe and de Haan will run further tests on McHugh, who is only too glad to be involved in the research. Since meeting Lythgoe, he says, he has come to terms with the intense bouts of creativity that at first terrified him. He has given a talk about his work at the Dana Centre in London — a venue dedicated to debates on science. He has also met other artists who have the same compulsions as him — without brain damage. "They experience the same thing, so I realized I wasn't alone," says McHugh. "Meeting them has been my life-saver."

Jim Giles is a reporter for *Nature* based in London.

People power against climate change

Local government and communities needn't wait for the big players to decide on action.

Sir—Your Editorial “Carbon impacts made visible” (*Nature* 429, 1; 2004) mentions two bottom-up opportunities that need to be seized if we are to reduce the causes of climate change. But you omitted another, possibly more significant, one that is already happening.

Since 1993, local governments in towns, cities and municipalities around the world have been signing up to their own Kyoto-like commitments, largely as part of the International Council for Local Environmental Initiatives programme, “Cities for Climate Protection” (www.iclei.org). At the last count, almost 600 had signed up. In the United Kingdom, 60 towns and cities, including Middlesbrough,

have signed up to a similar programme, under the Nottingham Declaration on Climate Change. Added together, these communities produce nearly 10% of the world's greenhouse gas emissions. So these community-level targets should not be ignored.

Local governments are well placed to get to the nub of the problem in a way that global treaties and carbon trading can never achieve. They can get the message across to the residents, shopkeepers and drivers of the world that efficient, cheap-to-run homes, fridges, cars and so on can improve their quality of life while helping a global cause. Also, people can take action locally regardless of national priorities:

witness the wealth of climate-friendly activities going on in many US cities despite the now infamous anti-Kyoto stance of the Bush administration (see *Earthed* 9–11, May 2004).

So in an age when the power of communities seems to be shrinking against the might of world governments and multinationals, it's refreshing to see that bottom-up local action still has an important part to play in tackling climate change. Power to the people indeed!

Jim Gillon

*Community Protection Service,
Middlesbrough Council, PO Box 65,
Vancouver House, Central Mews,
Gurney Street, Middlesbrough TS1 1QP, UK*

Ground rules for dealing with anthropomorphism

Sir—Clive Wynne misrepresented my published views in his Concepts essay “The perils of anthropomorphism” (*Nature* 428, 606; 2004). The concept of ‘critical anthropomorphism’ helps to establish ground rules for dealing with the inevitable anthropomorphic tendencies that we, as sentient human beings, confront in trying to understand the behaviour of other species.

Critical anthropomorphism involves not only careful replicable observation, but also knowledge of the natural history, ecology, and sensory and neural systems of animals as well. Such knowledge itself changes and has led me to reinterpret examples of animal behaviour involving, for example, perception and cooperation, in my own research. Problems can also arise when scientists fail to recognize inadvertent anthropomorphism in ‘objective’ studies; colleagues and I have documented erroneous conclusions in areas as diverse as communication, foraging, conservation planning and zoo design. As noted in your News Feature “True colours” (*Nature* 428, 596; 2004), “It's only when you begin to see the world as birds do — detecting light in the ultraviolet spectrum — that the full subtlety of their behaviour is revealed.”

Wynne asserts that I use critical anthropomorphism to argue that death-feigning behaviour in snakes is best understood by assuming that animals have conscious states. What we have shown is that hognose snakes monitor their visual environment and make adaptive decisions enabling escape behaviour while appearing completely unresponsive, and thus they are

not unaware or unconscious of their surroundings when in this state, as previous authors had assumed. This problem with the word “conscious” is why I have argued for use of the term “private experience” (derived from behaviourist B. F. Skinner's “private events”) and not “consciousness” in the study of animal behaviour. The issues here are controversial enough to make careful attention to language essential.

Gordon M. Burghardt

*Department of Psychology and Ecology &
Evolutionary Biology, University of Tennessee,
Knoxville, Tennessee 37996, USA*

Political interference not needed in Iraqi science

Sir—Your News story on the assassination of scientists in Iraq, “Iraqi killings prompt calls for US to evacuate weapons scientists” (*Nature* 429, 115; 2004), included the statement that, according to officials at the US Department of State, “scientists felt excluded from an attempt by a largely expatriate group of Iraqi researchers to form an Iraqi academy of science”.

In fact, the meeting to formally establish the Iraqi National Academy of Science, held at the Royal Society in London in November 2003 with the participation of the French and US national science academies, involved 12 scientists and engineers, eight of whom are — and for many difficult years have been — based in Iraq.

It is disappointing that such a carelessly inaccurate allegation should be attributable to the US Department of State. Indeed, there is a danger that it may be misinterpreted as a politically motivated attempt to undermine

the new Iraqi academy. Iraqi science needs an independent academy “promoting pure and applied science for the service of the people and the country, and reviving Iraqi talents for the good of humanity”, as its charter states. It does not need political interference from within or outside Iraq.

Robert May

*The Royal Society, 6–9 Carlton House Terrace,
London SW1Y 5AG, UK*

No joke when a technical fault fools readers

Sir—I was sent a message from *Nature* offering a free electronic version of the 1 April 2004 issue that could be downloaded. Although I carefully followed the instructions I was not able to get access to it online. When I tried again I was refused access because only one copy was allowed and it had a record of my previous abortive attempt. When I reported this to the *Nature* customer-support staff I was informed that the free issue was no longer available. Was the free copy ever available? Or was it all an April Fool joke?

Michael Heather

*Law School, Northumbria University,
Ellison Place, Newcastle upon Tyne NE1 8ST, UK*

The 1 April issue of *Nature*, available for a limited time as a digital edition, was downloaded by almost 6,000 people from Newsstand, the third-party provider. However, a small number of people encountered some technical problems, which are being addressed by Newsstand (enquiries to support@newsstand.com). *Nature* apologizes for any inconvenience —Publisher, *Nature*

The vehicle takes the wheel

Can our brains wrest control from our genes to put our own interests first?

The Robot's Rebellion: Finding Meaning in the Age of Darwin

by Keith E. Stanovich

University of Chicago Press: 2004. 374 pp.

\$27.50, £19.50

Valerie M. Chase

What does it mean to be rational in the modern world? *The Robot's Rebellion* mines evolutionary biology and cognitive science in search of an answer. Believing that "our folk psychology remains sealed off from evolutionary insights and neurophysiological facts," psychologist Keith Stanovich offers readers a sweeping tour of theory and research, advancing a programme of "cognitive reform" that puts human interests first. Who else, one may wonder, might benefit from our cognition?

An answer lurked within the theory of natural selection for more than a century before Richard Dawkins exposed it in *The Selfish Gene* (Oxford University Press, 1976). Living things, said Dawkins, are our genes' way of making copies of themselves. In this sense, all organisms — including humans — are "gigantic lumbering robots", or "vehicles", for confederations of genes. Whereas animals sometimes behave altruistically, genes are reliably selfish. Not only do they compete mercilessly with other genes, but they sacrifice their vehicles' interests when doing so benefits their representation in the gene pool.

This gene's-eye view implies that our cognitive systems owe their existence to the genes that built us. This includes the mechanisms underlying automatic cognition, such as sensory perception and associative learning — which Stanovich refers to as the autonomous set of systems (TASS), and which laypeople might call intuition — as well as higher cognitive processes such as moral reasoning and scientific inference. To borrow Stanovich's metaphor, humans live on a long genetic "leash". As we evolved, genes endowed us with an analytic cognitive system with powers of flexible, abstract representation and reasoning that enhanced our — and crucially their — ability to survive and reproduce. In doing so, they unwittingly opened the door to the robot's rebellion. Using our analytic system, we can override TASS responses in situations where they serve the genes at our expense, much as we currently use contraception to thwart our genes' interest in having copulation lead to pregnancy.

By making the point that cognition is optimized at the level of genes, not of individuals, Stanovich puts a fresh spin on the



familiar claim that people are sometimes woefully irrational. Selfish-gene theory also lends narrative structure to the book, which might otherwise have seemed a jumble of ideas. However, many of the self-defeating behaviours that Stanovich discusses, which include taunting disfigured people, not putting seat-belts on children and being susceptible to advertising, appear no more beneficial to our genes than to us. Our cognitive failures can often be blamed on the mismatch between the modern world and the ancestral environments in which our brains evolved, rather than on the conflict between the interests of our genes and their vehicles. Accordingly, Stanovich also draws on ideas from evolutionary psychology.

Although Stanovich emphasizes that learning can shape TASS, he maps it loosely to evolved mechanisms for solving adaptive problems such as recognizing faces, choosing a mate and reading the minds of others. As he points out, it performs such complex feats with a speed and accuracy that the serial, rule-based analytic system cannot touch. Stanovich is in awe of TASS, but warns us not to trust it blindly. Whether or not the human brain is "at war with itself", as he claims, his examples demonstrate that our intuitive cognition can prevent us from realizing today's personal and societal goals, such as maximizing investment returns and presuming innocence in criminal courts. How can these by-products of our generally

adaptive thinking be avoided?

Here Stanovich falters. One challenge for his idea of cognitive reform is defining rationality. Initially, he suggests that products of the analytic system such as logic, probability theory and utility maximization, all of which put a premium on internal consistency in choice, define what is rational. Later, he observes that inflexible adherence to these norms can undermine our interests by, for instance, causing us to violate social rules. Another challenge is figuring out how to monitor the output of TASS without overburdening the analytic system, which is slow and computationally expensive; second-guessing all our choices would be a full-time job. To make his proposal more tractable, Stanovich might have tried systematically identifying situations where the fit between our evolved minds and

our modern world is poor, and specifying how to look for and rectify blunders in each situation.

Would-be cognitive reformers receive more guidance from Stanovich concerning the rationality of their beliefs and desires. In the most thought-provoking part of the book, he argues for a definition of rationality that encompasses adherence to standards of truth and morality, as well as internal consistency. For this he takes up Dawkins' idea of memes — units of culturally transmitted information that, like genes, have no interest in their hosts except as tools for propagating themselves. Stanovich proposes, for instance, that we root out harmful memes by rigorously testing the premises behind any that encourage us to risk our physical safety, or that, like some religious and political ideologies, forbid us to evaluate them.

Sadly, evolutionary ideas have split cognitive science into pro and con camps. Like some in the con camp, Stanovich too often mistakes evolutionary psychology for a defence of the behavioural status quo. But with *The Robot's Rebellion*, he sets himself apart from unreflective thinkers on both sides of the divide by taking evolutionary accounts of cognition seriously, even as he urges us to improve on what evolution has wrought.

Valerie M. Chase is at the Institute of Psychology, University of Basel, Missionstrasse 60/62, Basel 4055, Switzerland.

D. GIFFORD/SPL

A chrestomathy on fishes

Darwin's Fishes: An Encyclopedia of Ichthyology, Ecology and Evolution

by Daniel Pauly

Cambridge University Press: 2004. 366 pp. £55, \$80

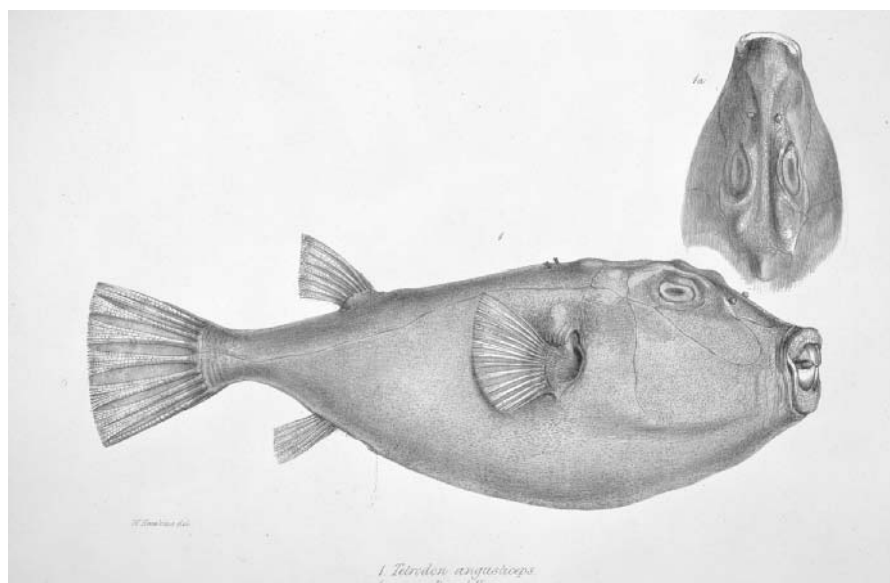
John C. Avise

Charles Darwin is renowned for his scientific breadth and prolific writing, as well as his depth of insight. On top of revolutionizing biological thinking (in *The Origin of Species* and *The Descent of Man*) and marine geology (in *The Structure and Distribution of Coral Reefs*), he wrote detailed treatises on creatures ranging from orchids and insectivorous plants to domesticated biotas, barnacles and earthworms. In addition, his name will forever be associated with some other taxonomic groups, notably Darwin's finches. During his career, Darwin wrote some 6,200,000 words of science, of which roughly 45,000 (nearly 1%) refer directly or indirectly to fishes. These have now been compiled, annotated, cross-referenced and elaborated on by Daniel Pauly in *Darwin's Fishes*, an encyclopedic treatment of these piscine works that the author professes as "a belated entry" on Darwin's behalf.

As a youth, Darwin was an avid angler, stating in correspondence that: "On the rainy days I go fishing, on the good ones Entomologizing." During the voyage of the *Beagle*, he collected and preserved in alcohol dozens of fish species, about which he recorded notes in *Fish in Spirits of Wine* (the piscine part of his *Catalogue*, or master list, of trip collections). Immediately after the voyage, Darwin successfully encouraged the British naturalist and clergyman Leonard Jenyns — who had previously turned down Captain Fitzroy's original offer to be the *Beagle's* naturalist — to complete the job of describing these fish specimens for science.

Later in life, Darwin incorporated his knowledge of fish into numerous works on topics that included the evolution of complex structures (such as electric organs in some rays and eels), sexual selection (in female-pregnant mollies and male-pregnant seahorses, for example) and the evolution of traits under artificial selection (as occurs in goldfish). Altogether, he refers to more than 250 fish species in his writings.

In *Darwin's Fishes*, Pauly describes everything that Darwin ever wrote on fish, the intent being to reveal how Darwin's world views were shaped by ichthyology, and vice versa. The book is encyclopedic not only in content but in style — its nearly 500 word entries (taxa or topics) are arranged alphabetically. A typical entry begins with a



Fish list: Darwin catalogued species such as *Sphoeroides angusticeps* during the voyage of the *Beagle*.

definition, incorporates relevant quotations or references from Darwin's writings, describes how his thoughts on the subject were influenced by studies on fish, and places his ideas in the contexts of both his own era and modern times. Pauly frequently voices his own welcome opinions also. For example, in his commentary on how Darwin (and later evolutionists) grappled with explaining the evolution of complex eyes, Pauly states: "This is strange, as the eyes are the organs for which intermediate steps that increase the fitness of their owners are not only easy to conceive, but occur in a wide range of animals." And with regard to Darwin's words on the behaviour of nesting wrasses, Pauly concludes: "One cannot but wonder whether the reason why these male wrasses work together so nicely with the females on their domestic arrangements is because, as protogynous hermaphrodites, they can draw on insights from their previous life as females."

Darwin's Fishes is not the type of book to be read from cover to cover but is wonderful fun to dip into, even randomly. It is an eclectic, scholarly reference work, rich in historical content and chock full of interesting quotes and insights from Darwin and Pauly alike. Pauly has an engaging writing style and is not afraid to use humour and self-deprecation. In the preface, for example, he short-circuits potential criticisms when he writes: "Many sections of this book read like laundry lists. I have attempted to cover this up, mainly through levity, the result being that this book will probably irritate serious scholars, but still bore students to tears." These concerns have elements of truth, but are considerably overstated. This book is indeed an all-but-the-kitchen-sink 'chrestomathy' (you can find that word as an entry in *Darwin's Fishes*), but it is a delightful one. ■

John C. Avise is in the Department of Genetics, University of Georgia, Athens, Georgia 30602, USA.

Infinite beauty

Fractals and Chaos: The Mandelbrot Set and Beyond

by Benoit B. Mandelbrot

Springer: 2004. 308 pp. \$49.95, £38.50, €49.95

Kenneth Falconer

It was just over 20 years ago that the Mandelbrot set took the world by storm. Pictures of extraordinary complexity and beauty appeared in scientific and glossy magazines, on the walls of art galleries and classrooms, on posters and even on tablemats. With the increasing availability of personal computers, drawing the Mandelbrot set became a standard exercise for those learning programming, and it was frequently an addiction for computer buffs, who were able to explore its intricacy by forever homing in on parts of the structure. Perhaps it is not surprising that such a simple procedure enabling almost anyone to produce an object of immense sophistication and attractiveness caught the public imagination.

The definition of the Mandelbrot set, denoted by M , is indeed extremely simple. Given a complex number c , start at the origin 0 and follow the trail of points obtained by repeatedly applying the transformation $f(z) = z^2 + c$, that is, the sequence 0, c , $c^2 + c$, $(c^2 + c)^2 + c$, ... If these points never go far away from the origin then c is in M , but if they wander off to infinity, c is not in M . This straightforward check allows one to scan across a region of the complex plane to determine the extent of M .

Crude pictures of M show a main cardioid surrounded by circular 'buds' of decreasing size. But more detailed investigation, pioneered by Benoit Mandelbrot in 1980,

Science in culture

The pen is mightier than the camera

Medical drawings bring a new dimension to the Royal Academy's Summer Exhibition in London.

Martin Kemp

Why would anyone involved in the sciences resort to drawing by hand? In this era of digital photography, which generates high-resolution images that can be manipulated at will, and of devices such as electron microscopes that can be set up to see and represent an object without fallible human intervention, why use such an apparently outmoded technique?

The answer is that the act of accurate drawing is still the best way to guarantee intense looking on the part of the artist, and, in terms of communication with the viewer, it provides a visual language of unprecedented focus, directness and dynamism. A drawing can transcend temporal limitations in a way that no other static image can. Data from different stages in a process or procedure can be overlaid yet remain clear, and anyone witnessing the making of the drawing can be taken step-by-step through what is being demonstrated.

Any artist who draws, or anyone who has sat for a drawn portrait, will not need telling that drawing involves the most intense visual concentration. It is a way of "learning how to see", to quote the first president of the Royal Academy, Joshua Reynolds. It is appropriate that when David Hockney and Allen Jones, former enfants terribles of the 1960s, were asked to organize a section devoted to drawing at this year's Royal Academy Summer Exhibition, they decided to include examples from outside Fine Art.

The drawings that have created the greatest stir — and won a prize — are by the heart surgeon Francis Wells of Papworth Hospital in Cambridge, UK. The example shown here illustrates the repair of a post-myocardial infarction ventricular septal



defect. Wells explains that it illustrates "the incision in the left ventricle to get access to the ventricular cavities, the relationship of the papillary muscles and the underside of the atrioventricular valves, the lesions involved, in this case a hole in the septum, the placement of the sutures for the introduction of the patch to close the defect and the patch in place. Finally, the sutures necessary to close the incision are also shown."

The circumstances of its creation explain the impulsive character of the drawn line. This picture was drawn after the operation, over a cup of coffee in the surgeon's room, to explain the procedure to visiting and trainee surgeons. The speed and weight of line and the selective hatching are integral to the act of emphatic communication, picking out exactly what is essential, filtering out all the visual

clutter that would be generated in a photograph or video. Even for someone who has not seen the drawing being made, the sense of the motion of the draughtsman's hand, in accordance with the flow of his ideas, is clear, just as it would be in a rapid sketch by Hockney or Jones. Indeed, Wells' sketch shares notable affinities with Hockney's drawings, especially those of the 1960s.

Wells himself is fully conscious of the special quality of the act of communicative drawing in time and place: "Having just completed the operation, drawing one's thoughts feels like an extension of the operation with the hand-eye interworking." This demonstrates

"the superiority of drawing over other means of visual record, as it is a process of communicating thought and ideas as composed in the mind without the constraints of any equipment limitations."

It is interesting to witness moves to reintroduce drawing into the teaching of sciences and technologies. As Wells says, drawing is "powerful in teaching and can enhance explanation — acting out the move ... through the use of line. I feel this is very like the cognitive process in the artist, who will look and look before committing a line to paper."

The drawings are on show at the Summer Exhibition of the Royal Academy of Arts in London until 16 August.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK, and co-director of Wallace Kemp Artakt.

reveals much, much more: the buds are all surrounded by smaller buds, which in turn support even smaller ones, and so on. Homing in on the boundary of M reveals a menagerie of multi-branched spirals, dragons and seahorses. Hairs of imperceptible fineness extend from the buds, holding along their lengths minute replicas of the entire Mandelbrot set.

Is the Mandelbrot set just a pretty curiosity? Far from it. It is a fundamental parameter set that encodes an enormous amount of information about nonlinear processes. First, the position of a complex number c relative to M tells us a great deal about the iteration of the quadratic mapping $f(z) = z^2 + c$. The (filled-in) 'Julia set' at c consists of those complex numbers z whose iterates under repeated application of $f(z) = z^2 + c$ never wander far from the origin. This Julia set comprises a single piece precisely when c lies in the Mandelbrot set. (Interestingly, this topological dichotomy was noted by Pierre

Fatou and Gaston Julia in 1918–19, but it was many years before its real significance and delicacy was appreciated.)

Much more than this, the exact position of c in M , such as the bud in which it lies, gives a very full description of how the iterates of $f(z)$ behave: for example, whether there are periodic cycles. Even more surprising is that, although defined in terms of the simplest of nonlinear maps, $f(z) = z^2 + c$, the Mandelbrot set is 'universal' in that it underlies the behaviour of very large classes of more complicated nonlinear mappings, the likes of which crop up throughout modern mathematics and its applications.

The emergence of the Mandelbrot set in 1980 led to a flurry of activity among mathematicians trying to understand its structure and significance, resulting in some of the most impressive advances in pure mathematics in recent years. In 1982, Adrien Douady and John Hubbard proved the (far from trivial) fact that M is connected, though

it is still unknown whether M is locally connected — can you travel between nearby points of M staying inside M without making too long a detour? In 1998, Mitsuhiro Shishikura showed that the boundary of M has fractal dimension 2, which means that it is just about as complicated as can be, though it is still not known whether this boundary has positive area.

This is the fourth volume of Mandelbrot's *Selecta*, comprising edited reprints of the author's papers. Largely from the 1980s, these include the series of seminal papers that revealed the magnificence and omnipresence of the Mandelbrot set, together with other papers related to the iteration of functions. Several sections provide an overview of the work along with its scientific and historical background. One chapter has been written specifically to help the non-expert appreciate the rest of the book.

Much of the material does not require particularly technical knowledge, so the

book should be accessible to a wide readership. It provides a fascinating insight into the author's journey of seeing and discovering as the early pictures of the Mandelbrot set started to reveal a whole new world. It gives a feeling for his philosophy and approach of experimental mathematics — an approach that has changed the way we think about mathematics and science. ■

Kenneth Falconer is professor of pure mathematics at the School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK.

Living machinery

Bionanotechnology: Lessons from Nature

by David S. Goodsell

Wiley-Liss: 2004. 337 pp. £47.50, \$79.50, €66.70

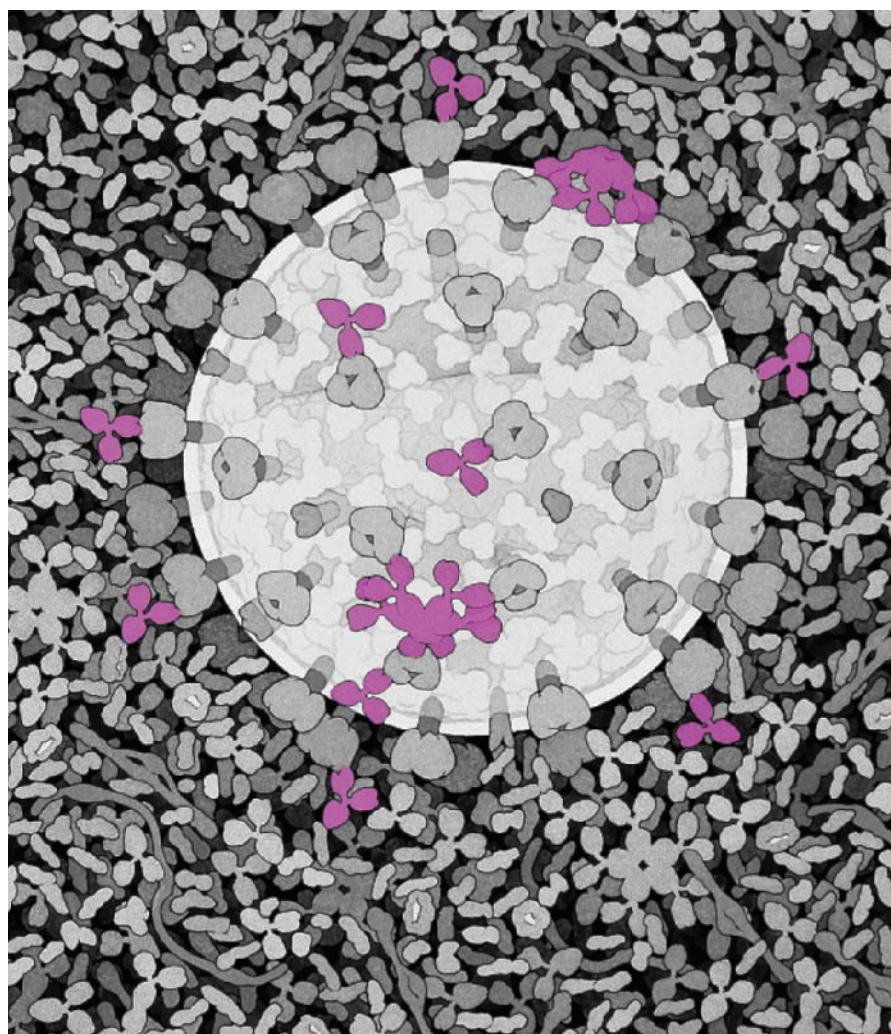
Christof M. Niemeyer

Nanotechnology is perfectly realized in biological systems. Cells are essentially biological assemblers that build thousands of custom-designed molecules and construct new assemblers. In *Bionanotechnology*, structural biologist David Goodsell describes what biology can teach us about engineering and manufacturing at the nanometre scale.

Goodsell kicks off with an introduction to the history of nanotechnology, which was pioneered by Richard Feynman and widely popularized by Eric Drexler's evocative idea of a self-replicating assembler building nanoscale devices atom by atom. He then takes a detailed look at the composition and structural principles of biomolecules harnessed in the cell, describing numerous bionanomachines in action, ranging from proteins, nucleic acids and membranes, to enzyme catalysis, the machinery of DNA transcription and translation, and biomolecular motors. The basic functional principles of nature's nanotechnology are elucidated, in particular the information-driven synthesis of biological molecules, the energetics and regulation of biological processes, and the traffic across membranes and signal transduction along them. The mechanistic aspects of biomaterials are also highlighted, for instance the interplay of myosin and actin filaments within the muscle sarcomere.

Throughout this fascinating journey, many fundamental principles are highlighted to introduce the reader to the unfamiliar world of nature's nanotechnology: the various bonding forces, thermal motion, the negligibility of gravity and inertia at the nanoscale, molecular recognition, self-assembly and biomolecular flexibility.

A survey of what is known about the general principles underpinning the structure and function of natural nanomachines is



Small wonder: antibodies (purple) are just one example of the way nature uses nanotechnology.

rounded off by an overview of the techniques that biotechnology has at its disposal to harness and modify this machinery — for example, recombinant DNA and protein technology, biomolecular structure determination, modelling and the tailoring of proteins and nucleic acids by directed evolution. The author then describes exciting applications of bionanotechnology that were the subject of published research at the time of writing in mid-2002. These include projects involving nanomedicine, self-assembly at various length scales, harnessing biomolecular motors, DNA computers, artificial life and biomolecular/bio(in)organic hybrid materials.

In the final chapter, Goodsell speculates about the future of bionanotechnology. He uses the best currently available data to relate three futuristic case studies: a “nanotube synthase” (an enzyme that would build nanotubes); a nanoscale assembler (a ribosome-like device that places molecular fragments in three-dimensional space with full positional control at the nanometre scale); and autonomous nanorobots that would inspect all the cells in the human body for mutations in the cancer-associated *p53* gene.

Bionanotechnology, being a synonym for nanobiotechnology, is a rapidly growing field that encompasses contributions from various disciplines, ranging from engineering and computational sciences to physics, chemistry and biology. Hence, there is an increasing need for a didactic textbook that provides a thorough introduction to biomolecular sciences and their impact on nanotechnology. Goodsell's book is the first to meet this demand.

Written in the style of an excellent biochemistry textbook, *Bionanotechnology* points the reader to general principles of the biological nanoworld, and thus provides readers with guidance on the design of their own devices and systems. The intention of this book is to invite further reading, so one should not expect too many technical details taken from the growing literature in this field. There are many graphic illustrations of bionanomachines, although it is a pity that these were not reproduced in full colour. I can highly recommend this book. I enjoyed reading every single page. ■

Christof M. Niemeyer is professor of chemistry at the University of Dortmund, Otto-Hahn Strasse 6, 44227 Dortmund, Germany.

A model for madness?

Dream consciousness: our understanding of the neurobiology of sleep offers insight into abnormalities in the waking brain.

Allan Hobson

Sigmund Freud and Carl Jung noticed the kinship of normal dreaming with psychoses of severe mental illness, such as schizophrenic and affective disorders. The presence in both of hallucinations and delusions, cognitive abnormalities and emotional intensifications had also been emphasized by earlier, more physiologically oriented psychiatrists, but the psychoanalysts soon eschewed neurobiology and attempted to explain both dreams and psychosis psychodynamically, without reference to the brain. A century of progress in neurobiology and in sleep and dream research now encourages a new look at this old question.

Recent advances in brain imaging, coupled with cellular and molecular neurobiology data, have given us a remarkably clear picture of the differences in brain activity between waking and activated states of sleep, such as rapid eye movement (REM), when intense dreaming most commonly occurs. During waking, consciousness is dominated by externally driven percepts, whereas internally generated images are rare. During REM sleep, the exact opposite rule applies — externally driven percepts are rare, whereas consciousness is dominated by internally generated images. Conversely, descriptions of thinking, which are common in reports of waking consciousness, are rare in reports of dreaming. When the mind is dominated by internal percepts, as it is in dreaming and psychotic states, the capacity to generate thoughts is greatly weakened. Rationality can thus, understandably, fall victim to a change in brain state.

To summarize the extensive data it is useful to look at the results of recent positron emission tomography (PET) scans and brain injury data from humans, and to compare them with earlier results of laboratory experiments in other mammals. The emphasis is on the distinct physiology of REM sleep, but it should be recognized that dreams associated with non-REM sleep are also created by a brain whose physiology has shifted strongly from that of waking towards that of REM.

One of the important causes of dream hallucinations is the general and widespread activation of the brain that occurs in conjunction with a blockade of external sensory inputs and motor outputs in REM sleep. In REM, selective activation of the cortical region responsible for visuospatial integration is seen.

During REM sleep, there is also selective activation of the amygdala and other parts of the limbic system. This is relevant for our



understanding of the heightened emotion — especially the feelings of anxiety, anger and elation — that so commonly dictate the development of a dream plot. An exaggeration of emotional activation is also common in mental illness, when it may contribute to impairment of rationality. An idea that helps to unite these findings is that emotion is itself an internally generated percept that powerfully influences conscious experience.

Human PET studies reveal a selective deactivation of the dorsolateral prefrontal cortex in REM state compared with waking. This brain region, which is assumed to be essential to working memory, directed attention and volition, remains inactive during REM. Subjects tell us that in this state they cannot use working memory effectively, are unable to sustain attention and do not will their dream actions. Selective frontal lobe deactivation has also been reported in PET studies of schizophrenia, and may account for the difficulties that these patients have in organizing their thoughts, integrating them with emotion, and translating them into appropriate actions.

Three aminergic neuronal systems — the noradrenergic, the serotonergic and the histaminergic — all shut down in REM. At the same time, the cholinergic system of the pontine tegmentum is activated. Thus, the entire brain is activated but aminergically demodulated except for dopamine (whose release continues at waking levels across all states of sleep). An abnormal sensitivity to dopamine is thought to mediate psychosis and its unmodulated action in REM may contribute to the madness of dreams. As nor-epinephrine is important for attention and serotonin is critical for learning and memory,

it is likely that the decreases in thinking during dreams are due in part to these remarkable and profound changes in brain chemistry.

As well as showing a decline in the capacity to think, formal state studies of dreaming also reveal a decrease in the capacity for orientation (times, places and persons change without notice) and a corresponding increase in the capacity to confabulate (giving the cognitive defects a coherence which is emotionally salient however illogical). Dream bizarreness and the confabulatory quality, coupled with the memory defects and the increase in visual hallucinations, suggests that dreaming may be a functional delirium, much like the psychosis associated with alcohol and drug abuse. Indeed, the delirium-like dreams of normal sleep are produced by changes in the same brain neuromodulatory systems that are affected by drugs en route to clinical delirium.

Now that we have this preliminary, but precise, model for how consciousness is altered by normal and abnormal changes in the state of the brain, we can conduct more hypothesis-driven studies of the brain–mind connection. For example, we can study regional-brain activation in the same animals from which we have learned so much at the cellular and molecular level. In those animals, we can also use drugs to alter a brain state experimentally and predict changes in regional activation pattern. In humans, we can use behavioral techniques as well as drugs to experimentally alter the conscious state of our human subjects; and we can measure predicted changes in regional brain activation with the real-time technique of functional magnetic resonance imaging.

To say that the mind–body problem is solved would be provocative and invite derision, but we have made important progress, and future directions of this approach are now clearly indicated. The mind–body problem can now be confronted by taking advantage of the natural changes in brain–mind state that contrast waking and dreaming. ■

Allan Hobson is in the Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA.

FURTHER READING

Hobson, J. A. *13 Dreams Freud Never Had: A New Mind Science* (Pi Press, New York, 2004).

Hobson, J. A. *et al. Behav. Brain Sci.* **23**, 793–842 (2000).

Snyder, S. H. *Drugs and the Brain* (Scientific American Books, New York, 1986).

Weinberger, D. R. & McClure, R. K. *Arch. Gen. Psych.* **59**, 553–558 (2002).

Solms, M. *The Neuropsychology of Dreams* (Lawrence Erlbaum Associates, New York, 1997).

As time glows by in bacteria

Carl Hirschie Johnson

Populations of cells can exhibit remarkably precise and stable circadian oscillations. But can single cells achieve such precision in the absence of intercellular communication? For cyanobacteria, it seems so.

Have you ever heard of a bacterium with jet lag? The idea might seem absurd; 20 years ago, experts in the field would have thought so too. Bacterial cells were believed to be just too 'simple' to have the biological clocks necessary to regulate daily (circadian) rhythms of gene expression, physiology and behaviour¹. But things have come a long way since then. We now know that cyanobacteria — also known as blue-green algae — have a robust circadian clock system that regulates the expression of all the genes in the organism. In fact, these bacteria are on the cutting edge of research into biological clocks, especially as regards structural studies of clock proteins and rigorous assessments of the evolutionary significance of circadian rhythms. And their utility looks set to continue: on page 81 of this issue, Mihalcescu and co-workers² report data that spotlight cyanobacteria as a new model system for the analysis of circadian rhythms in single cells.

It was already known from studies of non-bacterial (eukaryotic) organisms such as the unicellular species *Acetabularia*^{3,4} and *Paramecium*⁵, as well as of dispersed mammalian nerve cells⁶, that single cells can contain a functioning circadian clockwork. But individual cells of these types have rather 'sloppy' daily oscillations³⁻⁶, and circadian rhythms in populations of these cells tend to become dampened with time in constant conditions. This damping suggests that the cells' clocks have become desynchronized, because of noise and the variability in the oscillatory periods of individual cells in the population.

Nonetheless, it has long been recognized that the communication of phase information from one cell to another in a multicellular system could lead to the coupling of individually noisy cell oscillations to produce an emergent oscillating network that is precise⁷. An analogy would be the applause of human audiences, in which the synchronous clapping of individuals in a group is more precise than that of isolated individuals⁸. Alternatively, imagine a chorus line kicking in unison; in the absence of a musical beat, we would expect the precision of the kicking rhythm to be greater for the 'kick line' than for an isolated dancer, because of feedback along the kick line.

Indeed, the circadian rhythms of gene

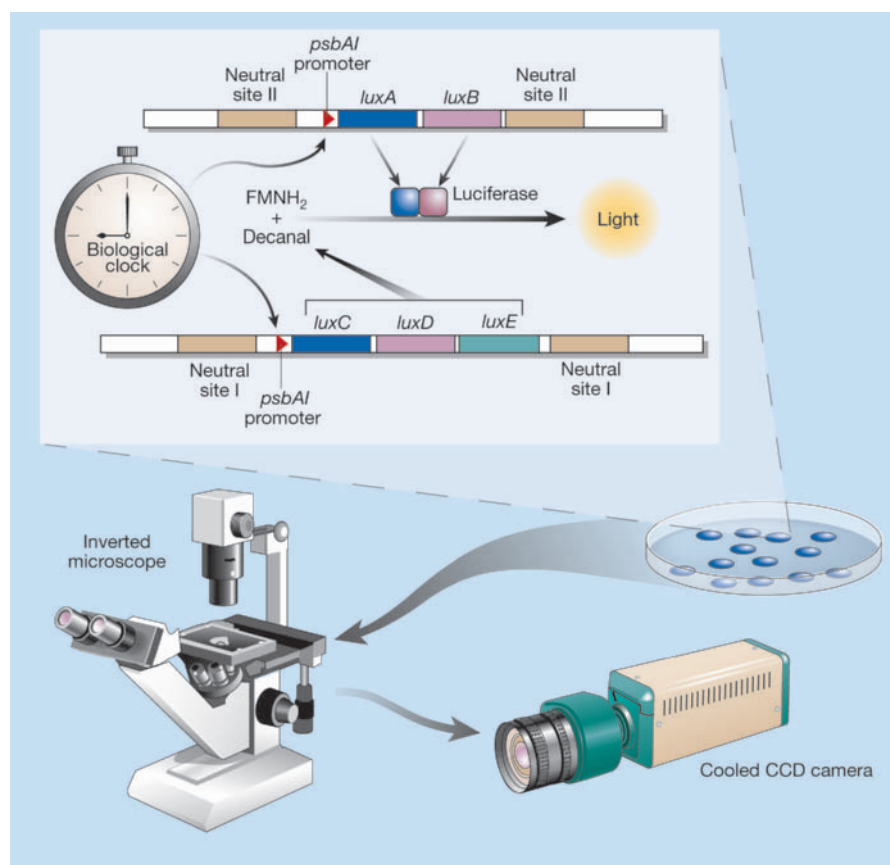


Figure 1 Monitoring circadian rhythms with self-luminescent cyanobacteria¹⁰. The bacteria are engineered to carry the *luxA–luxE* genes, under the control of the organism's own *psbAI* promoter region, which is in turn regulated by the cellular circadian clockwork. In a rhythmic fashion, the clock drives the expression of *luxA* and *luxB*, which together produce the luciferase enzyme. The *luxC*, *luxD* and *luxE* genes are also expressed, and produce enzymes that make the molecule decanal. The luciferase then catalyses a light-emitting reaction, using decanal and the bacterium's own FMNH₂ (reduced flavin mononucleotide) as substrates. Mihalcescu *et al.*² used this system with a sensitive microscopic imaging apparatus to analyse the luminescence from single isolated cells or cells in microcolonies.

expression in cyanobacterial populations exposed to constant conditions are impressively stable and precise¹⁹. My colleagues and I, who investigate cyanobacterial rhythms, thought it likely that the cells must be communicating with each other to achieve such precision. If so, the oscillations of isolated bacterial cells should be significantly more sloppy than those of cells in the population.

Enter Mihalcescu and co-workers², unfettered by our biases. These authors used a self-luminescent strain of cyanobacteria¹⁰ that glows when a gene regulatory region (the *psbAI* promoter region) is turned on by

the circadian clockwork (Fig. 1). They then monitored the luminescence rhythms of these cells in an automated, sensitive, microscopic imaging apparatus. Detecting the tiny pinpricks of light that are emitted by single bacterial cells is technically akin to detecting the dim glow of distant stars, and the cameras used for these two purposes are similar.

The results were dramatic: the authors found exquisite luminescence rhythms that are not markedly sloppier than those of populations. Moreover, as the bacteria divided, their daughter cells remained synchronized with each other remarkably well — cell division did not significantly perturb the

luminescence rhythms. So this circadian clockwork is an independent timing system that operates stably throughout the profound changes that cells undergo as a result of division, as had previously been found to be true for populations of cyanobacteria^{11,12}.

But how do the daughter cells remain so remarkably synchronized? And how can the cycle-to-cycle period be so precise? Is it a property of each individual bacterium, or does it result from the coupling of individually sloppy cell oscillations? Studies in eukaryotic algae supported the idea of a cell-to-cell communication of phase¹³. Mihalcescu and co-workers brought their technical prowess to bear on this question as it relates to cyanobacteria, by observing what happens when two microcolonies that start at different phases are merged with each other.

Specifically, they tracked the luminescence rhythms of each individual bacterium and its daughters during the merger. If the bacteria were communicating phase between each other, we might expect that the cells in the joined populations would reach a consensus. A single 12-hour dark pulse will rapidly synchronize all the cyanobacteria in a population — so these cells are indeed capable of rapid clock resetting. But what happens when out-of-phase colonies merge? The marvelous result is that a bacterium retains its specific phase for at least three days, even though it is touching out-of-phase cells.

Mihalcescu and co-workers' results do not exclude the possibility of weak coupling between the cells in a population. Another unresolved issue is whether the repeated dark exposures required for luminescence imaging (30 minutes every 2 hours) might have enhanced the inherent precision of the bacterial clock by frequency demultiplication. Nevertheless, the results certainly exclude the likelihood of any strong cell-to-cell communication of phase. Consequently, these 'simple' bacteria have circadian systems that are stable in the face of the cell division, intracellular noise and altered metabolic conditions that occur as a colony grows. In other words, cyanobacterial clocks can dance alone with high precision.

The biochemistry that underlies circadian clocks is remarkable. Not only is it temperature compensated and highly precise, with a time constant of 24 hours, we now know that it can be highly resistant to noise and is expressed at a cellular level, even in bacteria. A bacterium with jet lag? Not so absurd after all. ■

Carl Hirsch Johnson is in the Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235, USA.

e-mail: carl.h.johnson@vanderbilt.edu

1. Johnson, C. H. *Curr. Issues Mol. Biol.* **6**, 103–110 (2004).
2. Mihalcescu, I., Hsing, W. & Leibler, S. *Nature* **430**, 81–85 (2004).
3. Karakashian, M. W. & Schweiger, H. G. *Exp. Cell Res.* **97**, 366–377 (1976).

4. Woolum, J. C. *J. Biol. Rhythms* **6**, 129–136 (1991).
5. Miwa, I., Nagatoshii, H. & Horie, T. *J. Biol. Rhythms* **2**, 57–64 (1987).
6. Welsh, D. K., Logothetis, D. E., Meister, N. & Reppert, S. M. *Neuron* **14**, 697–706 (1995).
7. Enright, J. T. *Science* **209**, 1542–1545 (1980).
8. Neda, Z., Ravasz, E., Brechet, Y., Vicsek, T. & Barabasi, A.-L. *Nature* **403**, 849–850 (2000).
9. Kondo, T. *et al. Science* **266**, 1233–1236 (1994).
10. Katayama, M., Tsinoremas, N. F., Kondo, T. & Golden, S. S. *J. Bacteriol.* **181**, 3516–3524 (1999).
11. Mori, T., Binder, B. & Johnson, C. H. *Proc. Natl Acad. Sci. USA* **93**, 10183–10188 (1996).
12. Kondo, T. *et al. Science* **275**, 224–227 (1997).
13. Broda, H., Brugge, D., Homma, K. & Hastings, J. W. *Cell Biophys.* **8**, 47–67 (1985).

Extrasolar planets

Too close for comfort

Gibor Basri

Three gas-giant planets have now been found in puzzlingly close orbits around their stars. These 'very hot Jupiters' raise questions about planet-finding methods and our understanding of planetary systems.

Before 1995, the Solar System was our only known example of a system of planets orbiting a central star. We had convinced ourselves that its layout — small inner rocky planets and large outer gaseous planets — made sense on general principles. But then the first planet discovered outside the Solar System, the first extrasolar planet, fell completely outside the paradigm: it was a gas giant, similar to Jupiter in mass, but with an orbit eight times closer to its star than Mercury's is to the Sun¹. More of these 'hot Jupiters' (orbiting within a tenth of the Earth–Sun distance) were soon found, making it clear that the principles underlying the formation of planetary systems needed some revision. The detection since then of three 'very hot Jupiters' — two of which are now confirmed by Bouchy *et al.*² in *Astronomy and Astrophysics* — has strained the theory even further.

The first 'exoplanet' discovery, and the hundred that followed³, were all made using the Doppler-shift method. This indirect detection technique is based on the shift in the wavelength of a star's radiation as it 'wobbles' in a gravitational interaction with the orbiting planet. The Doppler technique is most sensitive to planets in close orbit around their stars, but no orbital periods shorter than about 2.5 days have been found with this technique. A possible explanation is that the disk of gas and dust from which the planets develop cannot itself approach the star too closely because of strong stellar magnetic fields⁴: the planets stop at the inner disk edge. Alternatively, the gas giants might be distended enough when they are very young to spill over onto the star if forced too close; eventually they achieve a minimum distance of separation⁵.

Another method for finding exoplanets exploits the dip in a star's brightness, as seen from Earth, when an orbiting planet crosses in front of it (as Venus did in front of the Sun last month). This is the transit method, and it is also most sensitive to exoplanets that are close to their stars, because the likelihood of a

planet aligning in front of its star, as seen from Earth, drops as the size of the orbit grows. Just as with the Doppler method, short orbital periods also make it more likely that a planet will be found — a factor that is even more crucial for the transit method, because there is no signal at all unless the planet is directly in front of the star. A much larger sample of stars must be observed to detect planets by transits than by the Doppler method.

The first detection of an exoplanet by the transit method was claimed only last year, in a survey of hundreds of thousands of distant stars⁶. The candidate planet is quite surprising, with an orbital period of 1.2 days. This is half the shortest period for a planet found using the Doppler-shift method⁷. Because there are many ways that pairs of stars can generate a false-positive signal for transit detections, transit candidates must be checked using the Doppler-shift method. Bouchy *et al.*² have now confirmed that two more transit candidates have convincing Doppler signals: their stars' spectral lines are shifted in exactly the expected manner. Both of these planets have shorter periods than any of the Doppler planets, at only 1.4 and 1.7 days, defining a new class of 'very hot Jupiters'. It is not clear how these planets could get so close to their star, nor how long they could last there — observations⁸ of the only transiting 'hot Jupiter' nearby show that it may be losing mass rather quickly. The theory of early spillover would also strongly apply to these new planets.

What is disturbing is that no 'hot Jupiters' were found in the transit survey, and no Doppler survey has found a 'very hot Jupiter'. Bouchy *et al.*² admit that they have not carefully analysed the sensitivity of their survey to planets at different orbital distances. They tentatively suggest that their detection rate of 'very hot Jupiters' might mean a general occurrence rate of one per 5,000 stars or so, which could be compatible with no detections among the roughly 3,000 Doppler-surveyed stars to date. They do not attempt, however, to explain why they found no 'hot

Jupiters'. The frequency of 'hot Jupiters' in the Doppler surveys is one per 125 stars, so their sensitivity to these would have to be down by a factor of 20–50 despite the fact that their orbital distances are only two to four times as large. This follows directly from the actual detection rates and stated planetary frequencies in the two types of survey.

We are left with an interesting mystery, which will require more observations and analysis to resolve. Either a new and insidious sort of false-positive for transit surveys has been found, or our theories of planetary systems need to be further expanded. In either case, the sensitivity of these surveys clearly needs to be better understood. Perhaps more observations will finally show that we need not distinguish between 'hot'

and 'very hot' Jupiters — planets may occur at all orbital distances. Meanwhile, we can continue to contemplate the production and fate of gas-giant planets roasting ever deeper in the outer atmospheres of their stars. ■

Gibor Basri is in the Department of Astronomy, University of California, Berkeley, California 94720, USA.

e-mail: basri@astro.berkeley.edu

1. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
2. Bouchy, F. *et al.* *Astron. Astrophys.* **421**, L13–L16 (2004).
3. Marcy, G. W. & Butler, R. P. *Annu. Rev. Astron. Astrophys.* **36**, 57–98 (1998).
4. Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. *Nature* **380**, 606–607 (1996).
5. Trilling, D. E. *et al.* *Astrophys. J.* **500**, 428–439 (1998).
6. Konacki, M., Torres, G., Jha, S. & Sasselov, D. D. *Nature* **421**, 507–509 (2003).
7. Udry, S. *et al.* *Astron. Astrophys.* **407**, 679–684 (2003).
8. Vidal-Madjar, A. *et al.* *Nature* **422**, 143–146 (2003).

Evolutionary genomics

Seeing double

André Goffeau

How do genomes evolve? Studies of numerous yeast species confirm the view that the duplication of genes, larger chromosomal segments and whole genomes are key mechanisms.

In a typical evolutionary tree, the yeast species stand on a branch apart from those of plants and animals. But the cells that make up these three types of organism have some crucial similarities, such as the presence of a nucleus and other subcellular structures. They are termed eukaryotic cells, in contrast to the simpler bacteria and archaea (prokaryotes). Because of these similarities, researchers have been able to use the rapidly reproducing and experimentally manageable yeast *Saccharomyces cerevisiae* to decipher some of the basic

properties of all eukaryotic cells. Now the use of other yeast species is opening up a new research front that seeks to identify the mechanisms behind genome evolution. The latest attacks on this front have come in the form of two papers published earlier this year in *Nature*¹ and *Science*², and one that appears now, on page 35 of this issue³.

The sequencing of whole genomes provides an interesting angle on the question of how species evolve, making it possible to look right down at the most fundamental — molecular — level. Researchers want to know

how, for instance, genes and genomes differ between species, and how these differences could have arisen while still allowing the organism to function. Only so much can be learned, however, by comparing the genomes of such evolutionarily distant organisms as, say, a yeast, a plant, a worm and humans, which were some of the first eukaryotes to have their genomes completely sequenced. So there is value in comparing different yeast species. Their genomes are relatively small and easy to dissect and their morphologies are similar, even though their evolutionary span is, at more than 1 billion years, longer than that of plants or animals.

After a pioneering screen of 13 partial yeast genome sequences⁴, several nearly complete sequences have become available over the past year or so^{1,2,5,6}. In two of the new papers, Kellis *et al.*¹ and Dietrich *et al.*² describe the sequences of the closely related yeasts *Kluyveromyces waltii* and *Ashbya gossypii*, respectively. These organisms are believed to have originated over 100 million years ago from an ancestor that is also shared by *S. cerevisiae*, and the authors' analyses confirm the suspicion^{7,8} that the *S. cerevisiae* genome was produced by wholesale duplication of the genome of that common ancestor.

How Kellis *et al.* and Dietrich *et al.* came to this conclusion is illustrated in Fig. 1. In brief, the authors found that many genes were homologous (indicating that they probably arose from the same ancestral gene) between *K. waltii* and *S. cerevisiae*¹, or between *A. gossypii* and *S. cerevisiae*². They also found that subsets of these genes were in the same order and orientation (they were 'contiguous') in the two genomes being compared. About 10% of these *A. gossypii* or *K. waltii* genes were found on both of two separate chromosomes in the *S. cerevisiae* genome. But, generally speaking, groups of contiguous *A. gossypii* or *K. waltii* genes were found on only one of these two *S. cerevisiae* chromosomes.

These results show that a genome duplication occurred in the *S. cerevisiae* lineage, after which some genes (about 10%) were retained in both locations. Some, however, were lost on one chromosome or the other. In some cases, duplicated genes presumably provided raw material for rapid evolution, as one gene of the pair could take on a new function, leaving the other to carry on as usual. But some duplicated genes remained similar, perhaps providing spare parts for the repair of genes encoding important proteins.

In the third paper, Dujon *et al.*³ describe their approach to the question of how genomes evolve. They present the complete genome sequences of four yeast species that are classified as hemiascomycetes — a broad phylum that evolved over a span of more than 500 million years. These species comprise the human pathogen *Candida glabrata*, the milk-loving *Kluyveromyces lactis*, the

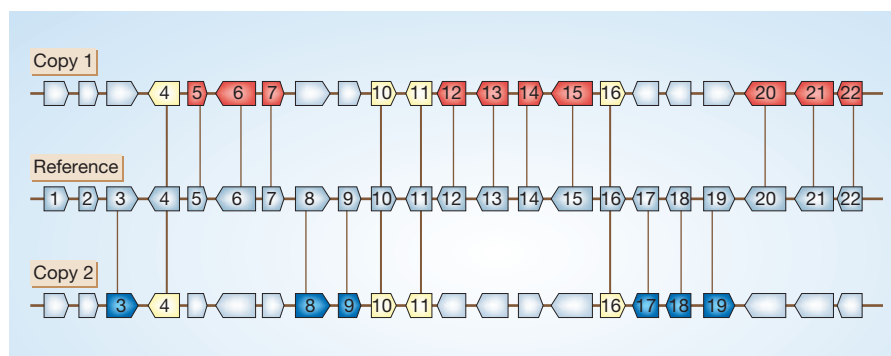


Figure 1 Detecting whole-genome duplications. It had long been suspected^{7,8} that the genome of the yeast *Saccharomyces cerevisiae* arose through the duplication of the genome of an ancestral yeast. Three new papers^{1–3} confirm this suspicion. The confirmation involved sequencing the genomes of yeasts such as *Kluyveromyces waltii*, *Ashbya gossypii*, *K. lactis* and *Candida glabrata* (which served as reference species) and comparing their gene sequences, and their order and orientation, with *S. cerevisiae*. The genes in grey are in the same order and orientation in the reference species and *S. cerevisiae*. Some of these genes, those in yellow, have relatives that are found in two different chromosomal locations (copies 1 and 2) in the *S. cerevisiae* genome. Some, however, have relatives in the same order and orientation only on copy 1 (red), or only on copy 2 (blue). The findings indicate that genome duplication occurred in the lineage that led to *S. cerevisiae*; some genes were then maintained, while many others were lost or diversified.

salt-tolerant *Debaryomyces hansenii* and the alkane-using *Yarrowia lipolytica*. Despite their similar morphologies and lifestyles, these species are, as Dujon *et al.* show, molecularly more diverse than an entire animal phylum, the chordates.

Together with *S. cerevisiae*, the four new sequences offer a collection of 30,028 proteins, of which 80% could be classified into about 4,700 families of homologues. The authors' analysis of these families shows that all five genomes are highly redundant, with *K. lactis* having the fewest duplicated genes, and *D. hansenii* and *Y. lipolytica* the most. *S. cerevisiae* and *C. glabrata* exhibit intermediate levels of redundancy.

Using the same methods as Kellis *et al.* and Dietrich *et al.*, Dujon *et al.* also found that the genomes of *S. cerevisiae* and *C. glabrata* were duplicated at some point in their evolutionary past, and that massive gene loss then ensued. No trace of whole-genome duplication was observed in the other three species.

So the gene redundancy in these organisms has to be explained by other mechanisms. These include tandem gene formation and segmental duplication, which seem to operate to some extent in all yeasts. Tandem repeats result from the head-to-tail duplication of one single gene, or, more rarely, of two or three adjacent genes. This might be repeated up to nine times, producing small arrays of tandem repeats. In contrast, segmental duplications replicate a large block of up to 250 contiguous genes, independently of other blocks or of the complete genome. The formation of segmental duplications has been reproduced experimentally in *S. cerevisiae*⁹ and inferred in other yeast species⁴.

Numerous cases of such duplications had been reported before, but their quantitative extent in given genomes had never been assessed. Dujon *et al.* found that *D. hansenii* is much richer in tandem gene repeats (247 arrays) than are the other yeasts they studied — there are less than 100 in *S. cerevisiae*, for instance. (However, tandem arrays seem to be constantly formed, and possibly rapidly resolved, in all species, and so might not always be recognized.) In addition, the authors identified eight duplicated blocks of genes in *K. lactis*, five in *D. hansenii* and two in *Y. lipolytica*. These blocks might have originated from independent segmental duplications. There must, however, also be other mechanisms because, in the highly redundant *Y. lipolytica* genome, most of the duplications do not correspond either to whole-genome duplications or to recognizable segmental or tandem duplications.

Together these data^{1–3} support the proposal, made by Susumu Ohno¹⁰, that the diversification of gene functions during evolution requires prior gene duplication. Early in the evolution of yeasts, it seems that major mechanisms for this were segmental

and tandem gene duplication. Moreover, a whole-genome duplication followed by extensive gene loss occurred in specific lineages, and has left visible relics in *S. cerevisiae*, in all *Saccharomyces* species *sensu stricto* and in *C. glabrata*.

More than 100 other yeast genome sequences are expected to become available in the near future and should make it possible for researchers to assess the importance of the various duplication mechanisms for each species. Comparative analyses of their genomes will also provide information on the still hazy processes of gene dispersion and loss, genome rearrangements and size constraints, the acquisition of species-specific genes, the effects of sex, and other

mechanisms — many of which should also be applicable to other eukaryotes. ■

André Goffeau is at the Institut des Sciences de la Vie, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium.
e-mail: goffeau@fysa.ucl.ac.be

1. Kellis, M., Birren, B. W. & Lander, E. S. *Nature* **428**, 617–624 (2004).
2. Dietrich, F. S. *et al.* *Science* **304**, 304–307 (2004).
3. Dujon, B. *et al.* *Nature* **430**, 35–44 (2004).
4. Souciet, J. L. *et al.* *FEBS Lett.* **487**, 3–12 (2000).
5. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. *Nature* **423**, 241–254 (2003).
6. Cliften, P. *et al.* *Science* **301**, 71–76 (2003).
7. Lalo, D., Stettler, S., Mariotte, S., Slonimski, P. P. & Thuriaux, P. *C. R. Acad. Sci. III* **316**, 367–373 (1993).
8. Wolfe, K. H. & Shields, D. C. *Nature* **387**, 708–713 (1997).
9. Koszul, R., Caburet, S., Dujon, B. & Fischer, G. *EMBO J.* **23**, 234–243 (2004).
10. Ohno, S. *Evolution by Gene Duplication* (Allen & Unwin, London, 1970).

Palaeoclimate

Message from the fish teeth

Bernhard Diekmann

An innovative analysis of isotope signatures in fish fossils will help in understanding past ocean circulation and its role in the climate system, and in testing models for climate reconstruction.

How did Earth's oceans behave during past episodes of long-term 'greenhouse' warmth? On page 65 of this issue¹, Deborah J. Thomas describes her investigations into the most recent episode of extreme global warmth, which occurred during the early Cenozoic era, between roughly 65 million and 40 million years ago, and reached its peaks at 55 million and 52–50 million years ago. Along with high levels of greenhouse gases in the atmosphere, a different pattern of ocean circulation, compared with that of today, is thought to have been a central determinant of the heat budget of the climate system. From her geochemical analysis of fossil-fish remains from muds of the central Pacific, Thomas concludes that ocean circulation was indeed different — but that the difference was not the same as that proposed in a prominent explanation for the global temperature regime during the early Cenozoic.

According to various lines of evidence — geological, palaeontological and geochemical — the early Cenozoic was characterized by mostly ice-free conditions, the occurrence at high latitudes of terrestrial floras and faunas requiring warm conditions, and the poleward displacements of coral-reef builders and other warm-water marine organisms^{2,3}. High-latitude sea-surface water and deep-ocean water both warmed up significantly, while — on early indications — sea-surface temperatures in the tropics were similar to or even lower than those of today⁴. From the subsequent refinement of data derived from geochemical signals in planktonic organisms, however, the evidence now favours the view

that tropical sea-surface temperatures were also significantly higher during the warm era⁵.

The greenhouse climate of the early Cenozoic is generally explained by high concentrations of atmospheric carbon dioxide, which may have been several times greater than the natural pre-industrial level of the modern post-ice-age period⁶, which started about 10,000 years ago. The huge amounts of CO₂ required are thought to have been produced by processes such as massive volcanism in the North Atlantic region, increased oxidation of carbon-bearing soils and rocks, or gas venting from the ocean floor^{3,6}. But the distribution of ocean temperatures at that time, with low latitudinal and vertical temperature gradients, cannot be solely explained by greenhouse gases⁷. Instead, it has been proposed that the greenhouse effect was accompanied by stronger poleward heat transport, in the oceans or atmosphere, that gave rise to polar warming. A prominent hypothesis invokes the sinking and poleward flow of warm and salty tropical waters in contributing to heat gain at high latitudes^{7,8}.

Such a pattern of ocean circulation would have differed from the modern mode of circulation, in which the formation of dense, cold and relatively salty deep-water masses takes place at the high latitudes of the North Atlantic and in the seas around Antarctica. These phenomena are connected with the global circulation, both at depth and at the surface, resulting in heat exchange between the hemispheres. In the early Cenozoic, however, not only was the temperature distribution of ocean waters different, but so was the plate-tectonic configuration of the continents. This

configuration, with closed seaways around Antarctica and an unimpeded circulation around the Equator, must have produced a different ocean-circulation pattern.

Thomas¹ used an innovative geochemical approach to trace past deep-water circulation in the equatorial Pacific. Her method is based on the analysis of neodymium isotopes in the teeth and bones of fossil fish. The traditional way of reconstructing water-mass movements involves analysing the carbon-isotope signals in the calcareous remains of plankton in seafloor sediments. The advantage of neodymium-isotope tracers is that they are particularly robust against post-depositional alteration⁹. The basic concept is that deep-water masses of different origin can be identified by the neodymium signal inherited from the solutions and particulates that enter the ocean following weathering on land; this signal becomes incorporated into the hard tissue of the local fish. The potential sites of deep-water formation in the Atlantic Ocean, the Antarctic Ocean and the North Pacific were bordered by landmasses with different crustal ages, and therefore with different neodymium-isotope ratios, allowing them to be distinguished from one another.

The highlight of Thomas's study is her compelling evidence that, in response to the warm climate during the early Cenozoic, formation of deep-water masses switched from the southern high latitudes to the North Pacific. Thomas concludes that this switch was not a consequence of the prevailing ocean temperatures during that time, but that it was possibly triggered by a change in the spatial pattern of freshwater precipitation and run-off from land. This change apparently resulted in a freshening of the surface waters in the Antarctic, making them less

dense, while those of the North Pacific became slightly saltier and dense enough to sink. When freshwater run-off from the continents is factored in, the existence of a 'downwelling cell' in the North Pacific is also predicted by model simulations of ocean circulation in the early Cenozoic¹⁰.

Thomas's findings clearly underscore the notion that ocean circulation is altered during times of global warmth. But her identification of high-latitude downwelling in the North Pacific argues against the often-proposed mechanism that, during the early Cenozoic greenhouse interval, heat was transported towards the poles by the spread of tropical waters. This result helps to verify the conclusions deduced from model simulations¹¹, which also do not support that hypothesis. An overall lesson, then, is that a combination of both approaches — evaluation of geological records and computer-based modelling — is essential for understanding ancient climate systems. ■

Bernhard Diekmann is at the Alfred Wegener Institute for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany. e-mail: bdiekmann@awi-potsdam.de

1. Thomas, D. J. *Nature* **430**, 65–68 (2004).
2. Prothero, D. R. *The Eocene-Oligocene Transition: Paradise Lost* (Columbia Univ. Press, 1994).
3. Thomas, E., Zachos, J. & Bralower, T. J. in *Warm Climates in Earth History* (eds Huber, B. T., MacLeod, K. G. & Wing, S. L.) 132–160 (Cambridge Univ. Press, 2000).
4. Zachos, J. C., Stott, L. D. & Lohmann, K. C. *Paleoceanography* **9**, 353–387 (1994).
5. Pearson, P. N. *et al.* *Nature* **413**, 481–487 (2001).
6. Pearson, P. N. & Palmer, M. R. *Nature* **406**, 695–699 (2000).
7. Sloan, L. C., Walker, J. C. G. & Moore, T. C. Jr *Paleoceanography* **10**, 347–356 (1995).
8. Lyle, M. *Paleoceanography* **12**, 161–167 (1997).
9. Martin, E. E. & Scher, H. D. *Earth Planet. Sci. Lett.* **220**, 25–39 (2004).
10. Bice, K. L., Barron, E. J. & Peterson, W. H. *Geology* **25**, 951–954 (1997).
11. Bice, K. L. & Marotzke, J. J. *Geophys. Res.* **106**, 11529–11542 (2001).

Neurobiology

Sleep on it

Ilana S. Hairston and Robert T. Knight

Is the function of sleep to replenish energy resources or to modify neural connections in the brain? Recordings of the brain's 'reverberating circuits' evident during sleep shed light on the question.

“With regard to sleep and waking, we must consider what they are: whether they are peculiar to soul or to body, or common to both; and if common, to what part of soul or body they appertain.” This question was posed by Aristotle¹ in 350 BC. Although the dualist view of body and soul is not part of modern neuroscience, we are still at a loss to understand the function of sleep. On page 78 of this issue², Huber *et al.* provide elegant results that favour the role of sleep in neural plasticity and memory consolidation.

Sleep is homeostatically regulated — that is, prolonged wakefulness results in a proportional increase in the tendency to fall asleep³. The hallmark of the need for sleep is an increase in the brain's slow-wave activity, as seen in electroencephalographic (EEG) recordings during non-rapid eye movement (NREM) sleep, which is proportional to the time spent awake. This has led to the formulation of the two-process model of sleep regulation: one process that reflects the dynamics of homeostatic events, expressed as modulation of slow-wave activity as a



100 YEARS AGO

Religion and Science: Some Suggestions for the Study of the Relations Between Them.

By P. N. Waggett, M.A. Pp. xii+174. (London: Longmans, Green and Co., 1904.)

It is pleasant to find a book which seeks to deal from the religious standpoint with the relations between religion and science... The author speaks, with possibly undue modesty, of his own opinions on the “domestic” issues that divide biologists. Holding, as he does, that “natural selection remains scientifically the most probable and philosophically the most welcome account of the adaptations of animal and vegetable life,” he is perhaps inclined to attach too much weight to the arguments that have been brought forward by various scientific authorities on the other side.

Buy English Acres. By C. F. Dowsett.

Pp. 224. (Published by the Author, Winklebury, Basingstoke.)

This is not a book in the ordinary sense. It is a collection of miscellaneous arguments, extracts from books, and biographical notes, all intended to prove that pleasure and profit may be derived from the purchase of English land. The absence of any attempt at coherence or sustained economic discussion is atoned for, so far as possible, by the author's great earnestness. Apart from that, the book has no serious qualities. From *Nature* 30 June 1904.

50 YEARS AGO

The Collected Papers of Peter J. W. Debye Pp. xxii+700. (New York: Interscience Publishers, Inc.; London: Interscience Publishers, Ltd., 1954.)

To celebrate the seventieth birthday of Prof. Peter Debye, this volume containing a selection of his classical papers has appeared. No one during the past forty years has made so many significant contributions to modern physical chemistry as has Prof. Debye: his name will always be associated with such topics as X-ray scattering, the theory of strong electrolytes, dipole moments and, during more recent years, with the scattering of light by colloidal solutions... While the book is an important one for purposes of reference, it is most valuable in that it gives us a glimpse of the methods of attack adopted by one who, trained as a mathematician, became in turn professor of theoretical then experimental physics and finally professor of chemistry, in which subject he received the Nobel Prize. From *Nature* 3 July 1954.



Figure 1 Why sleep? Huber *et al.*² find that during slow-wave sleep (tracing at bottom) people improve their ability to perform a task that they had learned before going to sleep. The results provide evidence that slow-wave activity has a key role in learning.

function of sleep–wake history; and another that reflects the circadian regulation of the sleep–wake cycle⁴. Sleep deprivation has serious and diverse biological consequences — from impaired cognitive function to perturbation of the immune, hormonal and autonomic nervous systems, and up to and including death.

Two prominent hypotheses have been proposed to explain the physiological function of sleep. One suggests that NREM sleep provides a period of low metabolic demand, during which a deficit in neural energy resources is replenished or intracellular ‘housekeeping chores’ are accomplished. Specifically, the metabolic demand during wakefulness is postulated to decrease the ATP:AMP ratio, resulting in high levels of extracellular adenosine, which in turn binds to adenosine (A1) receptors, inhibiting neuronal activity and lowering metabolic demand⁵. Put simply, high metabolic demand during wakefulness incurs an energy deficit, which is corrected during sleep.

A second hypothesis suggests that sleep has a key role in neural plasticity — that is, in maintaining appropriate connections between neurons through respectively reinforcing and eliminating significant and accidental connections between synapses. Owing to its traditional association with dreaming, REM sleep was first implicated in learning and neural plasticity, with findings — mainly from animal studies — showing a correlation between the amount of REM sleep and performance on a learned task⁶. More recently, NREM sleep, and specifically slow-wave activity, has also been shown to play a critical part in both developmental and learning-induced plasticity⁷.

Both hypotheses predict that local neural

activation due to specific cognitive demands will cause an increase in slow-wave activity during subsequent NREM sleep. It is important to note, however, that only the plasticity hypothesis predicts that such increased slow-wave activity would enhance performance during subsequent testing.

Huber *et al.*² provide results that favour the plasticity hypothesis. In their experiments, human subjects performed a complex task requiring hand–eye coordination before they went to sleep. This task typically involves the regions in the right parietal cortex of the brain⁸. Using a combination of techniques — an EEG system recording from 256 electrodes on a subject’s scalp and magnetic resonance imaging — Huber *et al.* found that, during ‘post-training’ sleep, slow-wave activity was increased only in the regions of the right parietal cortex thought to be involved in this task. The authors also demonstrated that sleep enhanced performance in a subsequent re-test on the task.

These findings complement previous observations, in both rats⁹ and humans¹⁰, that showed that stimulating events induce local changes in slow-wave activity, and that the amount of NREM sleep predicts performance improvement¹¹. But Huber and colleagues critically extend previous work by showing that the extent of the local increase in parietal slow-wave activity predicted performance enhancement in subsequent testing, thus supporting the view that slow-wave activity has a specific role in plasticity and learning.

The slow waves seen in EEG recordings are the result of synchronized oscillatory activity in cortical neurons. Work with *in vitro* preparations of neurons and with neural-network models has suggested that these slow oscillations are conducive to synaptic strengthening,

thereby contributing to memory consolidation^{12,13}. Support for this view has been proposed from rodent studies in which firing patterns of ‘place cells’ in the hippocampus, recorded in the awake animals, were found to be reactivated (‘replayed’) during subsequent NREM sleep¹⁴. Huber and colleagues’ finding, that local generation of slow-wave activity is associated with enhanced performance, provides proof of concept for these models in humans (Fig. 1).

The new work² offers evidence that neural plasticity occurs during sleep. But does this mean that plasticity is the function of sleep? If only part of the brain is undergoing synaptic strengthening, why is it necessary for all the rest of the brain and body to sleep? Synchronous neural activity for plasticity needs could be generated locally in the brain, without the broad behavioural and physiological changes that occur during sleep and that potentially expose the organism to life-threatening danger.

It has been proposed that the entire cortex of the brain experiences neural plasticity related to updating experiences of the world, and dependent on the previous day’s events¹⁵. If this were true, with appropriate testing one should be able to show that there are multiple regions of experience-dependent changes in slow-wave activity. But it remains to be seen how such regional changes would translate into the global dynamics determined by sleep–wake history. Conversely, it is possible that synaptic strengthening is an added benefit of slow-wave activity, which is actually driven by the need for metabolic recovery. Are these ‘reverberating circuits’ essential elements in the function of sleep, or are they a by-product of design for another purpose — are they, in fact, the spandrels of St Mark’s of sleep research. ■

Ilana S. Hairston is in the Department of Radiology, University of California at San Francisco, San Francisco, California 94143, USA.

e-mail: hairston@radiology.ucsf.edu

Robert T. Knight is in the Department of Psychology and the Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, California 94720, USA.

e-mail: rtknight@socrates.berkeley.edu

1. Aristotle *On the Soul* Vol. VIII, Parva Naturalia.
2. Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. *Nature* **430**, 78–81 (2004).
3. Carskadon, M. A. & Dement, W. *Percept. Motor Skills* **48**, 495–506 (1979).
4. Borbély, A. A. *Hum. Neurobiol.* **1**, 195–204 (1982).
5. Greene, R. W. & Siegel, J. M. *Neuromol. Med.* **5**, 59–68 (2004).
6. Smith, C. *Behav. Brain Res.* **78**, 49–56 (1996).
7. Benington, J. & Frank, M. G. *Prog. Neurobiol.* **69**, 71–101 (2003).
8. Cohen, Y. E. & Andersen, R. A. *Nature Rev. Neurosci.* **3**, 553–562 (2002).
9. Vyazovskiy, V., Borbély, A. & Tobler, I. *J. Sleep Res.* **9**, 367–371 (2000).
10. Kattler, H., Dijk, D. J. & Borbély, A. A. *J. Sleep Res.* **3**, 159–164 (1994).
11. Walker, M. P. *et al. Learn. Mem.* **10**, 275–284 (2003).
12. Sejnowski, T. J. & Destexhe, A. *Brain Res.* **886**, 208–223 (2000).
13. Steriade, M. & Timofeev, I. *Neuron* **37**, 563–576 (2003).
14. Lee, A. K. & Wilson, M. A. *Neuron* **36**, 1183–1194 (2002).
15. Kavanau, J. L. *Brain Res. Bull.* **42**, 245–264 (1997).

Astronomy

Where are Procyon's quakes?

Jørgen Christensen-Dalsgaard and Hans Kjeldsen

The acoustic waves that ripple through the Sun should exist in other stars too. But a search for these 'starquakes' in the nearby star Procyon has found no evidence of them.

More than a quarter of a century ago the Sun was found to be a variable star, the velocity of its surface gas varying with periods of around five minutes. A breakthrough came in 1979 with the first detection of large-scale oscillations on the solar surface^{1,2}. These oscillations, or 'starquakes', are acoustic waves that penetrate the solar core — and hence provide information about the sound speed and, less directly, about the temperature and composition of the deep solar interior. The oscillations are believed to be driven by acoustic noise generated by the vigorous convective motion of gas near the solar surface. Such motion dominates the energy transport in the outer parts of the Sun and of other stars with surface temperatures below 7,000 K.

Consequently, quakes similar to those of the Sun are expected in other stars as well, and could be a useful source of information about them. However, such 'asteroseismic' investigations are formidably difficult, owing to the small amplitude of the oscillations: the change they cause in the intensity of the Sun's radiation is at a level of only a few parts per million (p.p.m.). On page 51 of this issue, Matthews *et al.*³ report the first results from the Canadian MOST satellite, a space mission dedicated to asteroseismology. In these observations of the star Procyon (Fig. 1), there is no evidence of the expected acoustic oscillations — a telling indication of our limited understanding of the physics of Sun-like stars.

The amplitudes of the oscillations depend on the rate of convective-energy input and the rate of damping (which again largely depends on convective effects). The modelling⁴ of these processes is uncertain, but oscillations in stars hotter than the Sun are expected to have substantially higher amplitudes⁵. Procyon, one of the nearest Sun-like stars to Earth, is 1.5 times heavier than the Sun, but is more than seven times brighter; oscillation amplitudes of 18 p.p.m. have been predicted⁶. Also, the damping rates are predicted to vary relatively little with stellar parameters: in the Sun, the observed lifetime of the most strongly excited modes is 3–4 days, so something similar would be expected for Procyon.

In fact, there have already been indications of oscillations on Procyon^{7,8}. The changing radial velocity of the gas near the star's surface during quakes causes detectable shifts in the wavelength of the star's radiation (the Doppler effect). But the



Figure 1 Procyon in the winter sky. On the right is the easily recognizable constellation Orion, including the red star, Betelgeuse. Sirius, the brightest star in the night sky, is in the lower middle of the image. And in the upper left section is Procyon, the brightest star in the constellation Canis Minor.

measured amplitudes of the velocity oscillations turned out to be well below theoretical expectations.

Oscillations in stellar intensities have also been sought. These measurements are intrinsically simpler to make, but there are, however, complications. For ground-based telescopes, fluctuations in Earth's atmosphere must be taken into account. Achieving sufficient resolution requires observations over several days; yet, from a single observing site, continuous observations of typically no more than 10–11 hours per night are possible, introducing gaps in the data that greatly complicate the analysis. Data can be combined from several observatories, but this involves obvious organizational difficulties. Nevertheless, in a major effort⁹ to observe stars in a stellar cluster, involving several large telescopes, a detection sensitivity of 16–18 p.p.m. was achieved for the intensity oscillations. According to predictions, this should have been sufficient to pick up the oscillations in some of the stars observed, but no oscillations were detected.

To avoid the effects of Earth's atmosphere and to acquire a virtually uninterrupted data set, observations from space are the answer. The MOST satellite carries a 15-cm telescope, in principle allowing the instrument to reach a detection sensitivity of 3–4 p.p.m., significantly better than previous photometric observations. This relatively low-cost satellite (affectionately known as the

Hubble Space Telescope) was launched on 30 June last year, and initial tests showed that its capabilities, particularly the guiding system, met or exceeded expectations. It provided an important demonstration of the scientific potential of this class of microsatellite, and represents a major and eagerly awaited improvement in asteroseismic observational capabilities.

Procyon was an obvious first target, given its predicted properties and the existing observations of oscillations in radial velocity for that star. Indeed, it was expected that MOST would easily be able to provide a detailed analysis of the oscillation signal in Procyon. Yet no quakes have been detected³. How can this discrepancy be explained? The properties of the instrument or the star, or both, must be substantially different from expectations. And there is another complication to take into account — 'noise' from other fluctuations in the stellar atmosphere, particularly the direct effect of convection. Such effects are visible on the solar surface as 'granulation', with intensity fluctuations as high as 10%. Compared with the oscillation signal, this makes the star considerably noisier for intensity rather than velocity observations¹⁰.

As a radical solution, Matthews *et al.*³ raise the possibility that both the radial-velocity measurements and the MOST results are dominated by granulation noise, with no oscillation signal in fact being present in any existing data. As an alternative explanation, they propose that the properties of the oscillations are drastically different from the theoretical expectations.

We strongly support the latter possibility. The independent and largely consistent identification of oscillations in the velocity observations seems to be strong evidence that they do really exist. A preliminary analysis of the velocity data has also shown that the mode lifetime is probably less than two days — substantially lower than both the solar value (3–4 days) and the predictions. We doubt the authors' assertion that granulation dominates the noise, as this would require granulation noise in Procyon to exceed that in the Sun by a factor of five in amplitude, which is not supported by other observations or by hydrodynamic simulations of granulation. Thus we suspect that the MOST data might be dominated by non-stellar noise. This is supported by a preliminary simulation of the expected oscillation signal and granulation noise¹¹.

This conclusion should not overshadow the importance of the initial accomplishments of the MOST mission. Further analysis of the data is likely to elucidate the causes of the noise and may point to ways to compensate for it, allowing the Procyon signal to be detected. Also, the sensitivity and stability of the present measurements bode well for observations of oscillations in other stars that may have higher amplitudes or longer

A. FUJII/DAVID MALIN IMAGES

lifetimes. In the longer run, the experience gained with MOST on the nature of stellar pulsations, as well as on optimal techniques for space-based photometry, will be of crucial importance to coming asteroseismic missions — such as the French-based COROT mission and, one may hope, the far more ambitious Eddington mission originally selected for (but currently not included in) the programme of the European Space Agency. ■

Jørgen Christensen-Dalsgaard and Hans Kjeldsen are in the Department of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark. e-mails: jcd@phys.au.dk
hans@phys.au.dk.

1. Claverie, A., Isaak, G. R., McLeod, C. P., van der Raay, H. B. & Roca Cortes, T. *Nature* **282**, 591–594 (1979).
2. Grec, G., Fossat, E. & Pomerantz, M. *Nature* **288**, 541–544 (1980).
3. Matthews, J. M. *et al.* *Nature* **430**, 51–53 (2004).
4. Goldreich, P. & Keeley, D. A. *Astrophys. J.* **212**, 243–251 (1977).
5. Christensen-Dalsgaard, J. & Frandsen, S. *Solar Phys.* **82**, 469–486 (1983).
6. Kjeldsen, H. & Bedding, T. R. *Astron. Astrophys.* **293**, 87–106 (1995).
7. Brown, T. M., Gilliland, R. L., Noyes, R. W. & Ramsey, L. W. *Astrophys. J.* **368**, 599–609 (1991).
8. Martić, M., Lebrun, J.-C., Appourchaux, T. & Korzenik, S. G. *Astron. Astrophys.* **418**, 295–303 (2004).
9. Gilliland, R. L. *et al.* *Astron. J.* **106**, 2441–2476 (1993).
10. Harvey, J. W. in *Proc. IAU Symp. No. 123, Advances in Helio- and Asteroseismology* (eds Christensen-Dalsgaard, J. & Frandsen, S.) 497–511 (Reidel, Dordrecht, 1988).
11. <http://astro.phys.au.dk/~hans/procyon>

Cancer

Understanding the target

Michael R. Stratton and P. Andrew Futreal

The complicated responses of lung-cancer patients to a particular drug — gefitinib — are now less puzzling. Mutations in the target gene help to explain why the treatment works in some cases but not in others.

All cancers arise from mutations in key genes involved in cell proliferation, differentiation and death. Over the past 25 years, some 300 such mutated genes have been discovered¹, and the hope is that they, or the proteins they encode, will prove good targets in developing new anti-cancer therapies. One of the big successes has been the inhibition of certain mutated protein kinases by a drug called imatinib (marketed as Gleevec)². Protein kinases are enzymes that regulate the activity of their substrates by adding phosphate groups to them, and there are more than 500 encoded in the human genome³. Imatinib inhibits several protein kinases, and has proved remarkably effective in treating certain, comparatively uncommon, cancers in which these genes are mutated and activated². These discoveries have transformed cancer research, and an intensive search is on for inhibitors of other protein kinases mutated in human cancer.

It now seems that inhibition of a mutated and activated protein kinase may also be an effective way to treat a common cancer. Writing respectively in *Science* and *The New England Journal of Medicine*, Paez *et al.*⁴ and Lynch *et al.*⁵ report mutations in the epidermal growth factor receptor gene (*EGFR*) in a subset of lung cancers that make the disease more responsive to treatment. The *EGFR* protein is a kinase that spans the cell membrane. When bound to its ligand, epidermal growth factor, it is activated and triggers cell proliferation through a signalling cascade. The *EGFR* mutations found in lung cancers cause exaggerated and extended activation of the kinase in response to epidermal growth factor⁵. It is therefore highly likely

that these mutations are involved in causing the lung cancers in which they are found.

Clinical trials of the *EGFR* inhibitor gefitinib (marketed as Iressa) in lung cancer were originally based on the rationale that many such cancers have high levels of *EGFR* protein, which might be a factor in cancer development. In early trials of gefitinib alone, a few cancers responded and on this basis it was licensed. Subsequently, however, in two large randomized trials in which the drug was used in combination with others, no effect was seen⁶.

In these trials there was little correlation between the responses of individual cancers to gefitinib and levels of *EGFR* protein in the lung cancer. But with the results of Paez *et al.*⁴ and Lynch *et al.*⁵, it now seems that most lung cancers that respond to gefitinib have an activating *EGFR* mutation. In cancers that do not respond to gefitinib, the frequency of such mutations is much lower; in fact, the mutated *EGFR* found in lung cancers is more sensitive to inhibition by gefitinib than normal *EGFR*. The new studies provide evidence that the responses to gefitinib in early clinical trials were not flukes; that the drug probably works through *EGFR*, its presumed target (not a foregone conclusion because many protein-kinase inhibitors act against several targets); and that its therapeutic effect depends on the presence of activating mutations in the target protein.

Trials of inhibitors conducted exclusively on lung cancers with *EGFR* mutations should now clarify the overall benefits of gefitinib and similar drugs. Many responses are partial, and the proportion of cancers with *EGFR* mutations that do not respond is unknown. Moreover, the precedent of imatinib suggests

that cancers with different *EGFR* mutations may respond differently to inhibitors⁷. Finally, in cancers that initially respond and subsequently recur, the *EGFR* gene will be the first place to look for additional mutations that have conferred drug resistance.

How will these observations be translated into clinical practice? *EGFR* mutations are more common in women who have never smoked. Overall, however, fewer than 10% of lung adenocarcinomas in patients in the United States have an *EGFR* mutation (although the incidence may be higher than 30% in Japan). So, in the United States at least, it is probably not medically or economically reasonable to give gefitinib to all patients with lung adenocarcinoma. Instead, it will be necessary to identify the cancers with *EGFR* mutations and treat those patients. Some therapies directed against specific molecular targets are already administered according to the status of the target in the cancer (for example, tamoxifen in oestrogen-receptor-positive breast cancer, and trastuzumab in breast cancers with amplification of the *ERBB2* gene). But the newly revealed sensitivity to inhibitors conferred by *EGFR* mutations in lung cancer may finally usher in the era of personalized treatments based on the DNA sequence of the cancer cell.

For patients whose lung cancers have *EGFR* mutations, the results reported by Paez *et al.*⁴ and Lynch *et al.*⁵ will come as good news. For health providers, the prospect of reducing the cost burden of a new drug by limiting treatment to those who are likely to respond may be tempered by the large numbers of diagnostic tests required to identify susceptible cancers. For drug developers, the message is that cancer is a very diverse genetic disease. There are probably several more mutated protein kinases that are tractable drug targets. Like *EGFR*, many may be involved in only a minor fraction of any cancer type.

Is this kind of target commercially attractive? The thinking behind the development of *EGFR* inhibitors as cancer treatments was that they might be appropriate for a substantial proportion of cases of a common disease. The new studies certainly add weight to the evidence that gefitinib works in lung cancer. The irony is that if the results had been known in advance, they might have dampened enthusiasm for its development. ■

Michael R. Stratton and P. Andrew Futreal are in the Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK.

e-mails: mrs@sanger.ac.uk

paf@sanger.ac.uk

1. Futreal, P. A. *et al.* *Nature Rev. Cancer* **4**, 177–183 (2004).
2. Sawyers, C. L. *Genes Dev.* **17**, 2998–3010 (2003).
3. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. *Science* **298**, 1912–1934 (2002).
4. Paez, J. G. *et al.* *Science* **304**, 1497–1500 (2004).
5. Lynch, T. J. *et al.* *New Engl. J. Med.* **350**, 2129–2139 (2004).
6. Cohen, M. H. *et al.* *Clin. Cancer Res.* **10**, 1212–1218 (2004).
7. Heinrich, M. C. *J. Clin. Oncol.* **21**, 4342–4349 (2003).

Archaeology

Magnetic murals in Central America

Geophys. Res. Lett. **31**, doi:10.1029/2004GL020065 (2004)

The history of Earth's magnetic field over the past few millennia is written on the walls of archaeological sites in Central America, according to Avto Goguitchaichvili and colleagues, who have measured the magnetism of pre-Columbian mural paintings in central Mexico.

Palaeomagnetic records are generally reconstructed by looking at ancient magnetic rocks, which 'freeze in' the prevailing orientation of the geomagnetic field when they solidify from lavas. But relatively recent records are hard to find. Goguitchaichvili *et al.* point out that Mesoamerican murals retain an imprint of the planet's magnetic field at the time they were painted, because the grains of iron oxide used as a red pigment were free to orient themselves before the paint dried.

The authors have taken pigment samples from wall paintings spanning the period of the Classic and early post-Classic Mesoamerican cultures (about AD 200–1200). The samples all show remanent magnetization, mostly due to the presence of small amounts of magnetites (ferrimagnetic) and haematites (antiferromagnetic). Such data can be used to document changes in the direction and intensity of the geomagnetic field. Conversely, if these changes have already been well characterized for a particular region, the magnetic record could be used to date the murals, at a fraction of the cost of radiometric methods.

Philip Ball

Astronomy

Cool measurements

Astron. Astrophys. doi:10.1051/0004-6361:20040551 (2004)

Astronomers have made the first-ever measurement of the mass of an ultra-cool star and its companion brown dwarf. Stars generally create most of their light by fusing hydrogen in their cores, providing a direct relationship between their luminosity and their mass. But ultra-cool stars do not fuse hydrogen, and their mass–luminosity relationship is not a reliable way to measure their mass.

H. Bouy *et al.* now report the first step towards calibrating a separate mass–luminosity curve for these objects, which occupy a twilight zone between being stars and gas-giant planets. They collated four years of observations from various sources — the Hubble Space Telescope, the Very Large Telescope, the W. M. Keck Observatory and the Gemini

North Telescope — to determine the dynamical mass and orbital period of a binary system that lies about 40 light years from Earth.

They report that the heavier star has about 8.5% of the mass of our Sun, whereas its brown dwarf companion has only 6.6% of the solar mass. Both objects have a surface temperature of about 1,500 °C, more than three times cooler than the Sun. Bouy *et al.* estimate that the stars are relatively young, at between 500 and 1,000 million years old, and that they orbit one another every ten years or so.

Mark Peplow

Ecology

Waspish behaviour

Behav. Ecol. **15**, 661–665 (2004)

The butterfly *Melitaea cinxia* is parasitized by the wasp *Hyposoter horticola*, even though the host is susceptible for only a few hours each year. Saskya van Nouhuys and Johanna Ehrnsten conclude that the wasp achieves this feat by memorizing the locations of host eggs in advance. It then returns just before the butterfly larvae hatch, when the eggshells are weak enough for the wasp to penetrate and lay its own eggs inside (pictured).

Van Nouhuys and Ehrnsten made clusters of eggs available to the wasps for either two weeks before hatching or on the day of hatching. They found that the wasps parasitized a much greater proportion of the eggs that they had been able to identify on reconnaissance trips.

The authors point out that, as egg clusters are often hundreds of metres apart, it is unlikely that the wasps return to previously deposited chemical markers: known markers are short-range only. Instead, they think that the wasps are learning the spatial locations of the eggs. This would be a remarkable accomplishment — one that allows the parasite to rely on a single host species that is available for only a thousandth of that species' lifespan.

Patrick Stevenson



Neuroscience

Cocaine turn-off

Proc. Natl Acad. Sci. USA
doi:10.1073/pnas.0403795101 (2004)

What is the best way to tackle cocaine addiction? Many think that removing the psychological effects — the main driving force for addiction — might make addicts less inclined to take the drug.

M. Rocio A. Carrera and colleagues set about mopping up cocaine molecules in the brains of rats. The drug can be removed from the rest of the bloodstream by using cocaine-seeking antibodies, but the blood–brain barrier makes the brain a difficult target.

Carrera *et al.* devised a novel carrier molecule for the cocaine antibodies: a virus that can enter the brain. The virus was modified to carry genes for the antibodies and was made safe by having harmful gene sequences removed.

The rats received the carrier virus by nasal injection for three days. For the next few days, this group and an untreated group were given a dose of cocaine.

The untreated group displayed characteristic behaviour associated with taking cocaine, such as sniffing and moving backwards and forwards, whereas the treated group did not. Carrera *et al.* conclude that the antibodies had sequestered the cocaine molecules, blocking their physical and, probably, mental effects.

Laura Nelson

Metallurgy

A glass act

Phys. Rev. Lett. **92**, 245503 (2004)

Glassy steel promises to be a better engineering material than regular crystalline steel for certain applications — it can be harder, stronger, more corrosion-resistant and non-magnetic. But making it is not easy: previously, rods no more than 4 mm in diameter had been produced. Z. P. Lu *et al.* have found a recipe for making ingots of amorphous steel up to 12 mm in diameter, which should allow structural engineering components to be made using standard melting and casting techniques.

The group found that steel's ability to form a glass is enhanced by adding yttrium, together with small amounts of chromium, molybdenum, manganese and boron. Without yttrium, the alloy tends to crystallize; but mixtures with about 1.5% yttrium seem to be fully amorphous. The yttrium stabilizes the liquid phase relative to the crystal, and also slows the growth of iron carbide crystals — it makes an amorphous steel both thermodynamically and kinetically more favourable. The resulting material is twice as hard as standard steels, and is also harder and stronger than other (more expensive) amorphous alloys based on palladium and zirconium.

Philip Ball

Pterosaurs as part of a spinosaur diet

A rare fossilized action snapshot captures a mortal tussle with a hungry predator.

A remarkable specimen has been discovered of an Early Cretaceous pterosaur that has a tooth embedded in one of its cervical vertebrae: the tooth has been identified as one from a spinosaurid theropod dinosaur. This fossil is direct evidence that spinosaurs included items other than fish in their diet.

The diet of spinosaurid theropods such as *Baryonyx* and *Spinosaurus* has been the subject of speculation^{1–3}. Their peculiar elongated jaws, sinuous tooth row and crocodile-like teeth suggest that they were piscivores, a habit supported by the discovery of *Lepidotes* scales etched by gastric juices within the body cavity of the type specimen of *Baryonyx walkeri*, from the Wealden of Britain¹. However, bones of a juvenile *Iguanodon* were also associated with the specimen, suggesting that *Baryonyx* might not have been exclusively a fish-eater. *Baryonyx* may also have been a scavenger⁴. Still, until now there has been no direct evidence to indicate what the diet of other spinosaurs could have been.

The specimen we describe here (Fig. 1; Wyoming Dinosaur Center, WDC-SFB-001a,b,c), from the Early Cretaceous Santana Formation of northeastern Brazil, indicates that spinosaurs occasionally fed on pterosaurs, possibly during the course of scavenging activities. The specimen is part of a group of three pterosaur vertebrae enclosed in a nodule from the Romualdo Member of the Santana Formation⁵. The vertebrae are short and square-shaped in dorsal view, and are identified as cervical vertebrae⁶ numbers four to six of an ornithocheirid pterosaur, a

family that occurs frequently in the Santana Formation⁷. Their size indicates that they belonged to a creature with a wingspan of about 3.30 metres.

During the course of acid preparation of the nodule, we discovered that the anteriormost vertebra is perforated on the right side by a broken conical tooth (Fig. 1a). The tooth penetrates the vertebral centrum at the level of the anterior end of the pleurocoel, crosses the neural canal and damages the base of the neural arch, with its apex embedded in the left prezygapophysis (Fig. 1b). The thin, smooth enamel, straight crown, slightly compressed cross-section and lack of serrations on the lateral carinae show that the tooth is from a spinosaurid theropod.

Two spinosaurid taxa, *Irritator challengeri* and *Angaturama limai*, have been described from the Santana Formation^{3,8} (it is generally agreed that the latter is a junior synonym of the former). The tooth in the pterosaur vertebra is similar in every respect to the spinosaurid teeth previously described from the Santana Formation^{3,8}.

The specimen is a rare instance of fossil 'behaviour': the tooth reached its present position when a spinosaurid bit into the neck of a pterosaur. As revealed by the presence of three associated cervical vertebrae, the neck must have been fresh and articulated when it was bitten; it was not swallowed and digested, because the vertebrae remain in articulation and lack evidence of etching by gastric juices. Despite their thin walls and hollow construction, the vertebrae must have been very resistant for the dinosaur tooth to break as it was biting them.

In what is the only previously known instance of dinosaur predation on a pterosaur — an azhdarchid tibia bitten by a troodontid from the Late Cretaceous of Alberta⁹ — a broken theropod tooth is similarly associated with a pterosaur bone. In the case described here, the pterosaur may have been caught on the wing or on the ground; although trackways reveal an unexpectedly high terrestrial velocity in some pterosaurs¹⁰, creatures with a large wingspan may have been vulnerable on the ground. But these options

seem relatively unlikely and the scavenging of a pterosaur carcass on the shore of the lagoon in which the Santana Formation was deposited may be a more plausible explanation. Although the dinosaur did not swallow that part of the neck, this rare evidence of predation or scavenging shows that pterosaurs were part of the diet of spinosaurs, which we conclude were not strictly piscivorous.

Eric Buffetaut*, David Martill†, François Escuillié‡

*CNRS, 16 cour du Liégat, 75013 Paris, France

e-mail: eric.buffetaut@wanadoo.fr

†School of Earth and Environmental Sciences, University of Portsmouth, Portsmouth PO1 3QL, UK

‡Eldonia, 03800 Gannat, France

1. Charig, A. J. & Milner, A. C. *Bull. Nat. Hist. Mus. (Geol.)* **53**, 11–70 (1997).
2. Holz, T. R. *Science* **282**, 1276–1277 (1998).
3. Sues, H.-D., Frey, E., Martill, D. M. & Scott, D. M. *J. Vert. Paleontol.* **22**, 535–547 (2002).
4. Kitchener, A. *Nature* **325**, 114 (1987).
5. Martill, D. M. *Fossils of the Santana and Crato Formations, Brazil* (Palaeontological Association, London, 1993).
6. Wellnhofer, P. *Palaeontographica* **A215**, 43–101 (1991).
7. Kellner, A. W. A. & Tomida, Y. *Natl Sci. Mus. Monogr.* **17**, 1–135 (2000).
8. Kellner, A. W. A. & Campos, D. de A. *Neues Jb. Geol. Palaeontol. Abh.* **199**, 151–166 (1996).
9. Currie, P. J. & Jacobsen, A. R. *Can. J. Earth Sci.* **32**, 922–925 (1995).
10. Mazin, J. M., Billon-Bruyat, J. P., Hantzpergue, P. & Lafaure, G. in *Evolution and Palaeobiology of Pterosaurs* (eds Buffetaut, E. & Mazin, J. M.) 283–296 (Geol. Soc. Lond. Spec. Publ. **217**, London, 2003).

Competing financial interests: declared none.

brief communications arising online

▶ www.nature.com/bca

Biodiversity conservation: Uncertainty in predictions of extinction risk

W. Thuiller, M. B. Araújo, R. G. Pearson, R. J. Whittaker, L. Brotons, S. Lavorel (doi:10.1038/nature02716)

Biodiversity conservation: Effects of changes in climate and land use

L. B. Buckley, J. Roughgarden (doi:10.1038/nature02717)

Biodiversity conservation: Climate change and extinction risk

J. Harte, A. Ostling, J. L. Green, A. Kinzig (doi:10.1038/nature02718)

Biodiversity conservation: Reply

C. D. Thomas, S. E. Williams, A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. Ferreira de Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A. S. van Jaarsveld, G. F. Midgley, L. Miles, M. A. Ortega-Huerta, A. T. Peterson, O. L. Phillips (doi:10.1038/nature02719)

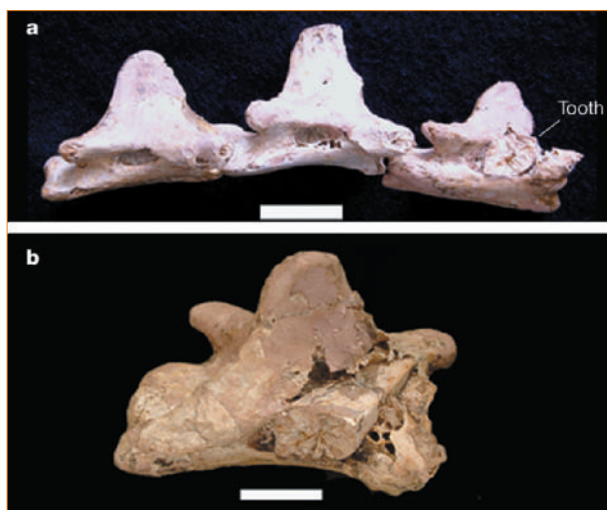


Figure 1 Cervical vertebrae of a pterosaur, with a dinosaur tooth embedded in one. **a**, Group of three cervical vertebrae (WDC-SFB-001a,b,c) of an ornithocheirid pterosaur from the Early Cretaceous Santana Formation of northeastern Brazil; right lateral view. The anteriormost vertebra is perforated by a spinosaur tooth. Scale bar: 20 mm. **b**, Close-up of pterosaur vertebra perforated by the spinosaur tooth, in dorsolateral view. Scale bar: 10 mm.

Biodiversity conservation

Uncertainty in predictions of extinction risk

Arising from: C. D. Thomas *et al.* *Nature* **427**, 145–148 (2004)

Thomas *et al.*¹ model species-distribution responses to a range of climate-warming scenarios and use a novel application of the species–area relationship to estimate that 15–37% of modelled species in various regions of the world will be committed to extinction by 2050. Although we acknowledge the efforts that they make to measure the uncertainties associated with different climate scenarios, species' dispersal abilities and *z* values (predictions ranged from 5.6% to 78.6% extinctions), we find that two additional sources of uncertainty may substantially increase the variability in predictions.

First, the study by Thomas *et al.* is based on projections of species-range shifts from a variety of niche-based models supplied by different contributors using different modelling methods. For instance, generalized linear models were used to model plants in Europe, whereas generalized additive models were used for *Protea* species in South Africa, and genetic algorithms for taxa in Mexico. Although niche-based models are all based on the same principle, they use a variety of assumptions, algorithms and parameterizations. Therefore, combining assessments from different models is likely to introduce further unquantified model effects.

To illustrate this, we fitted four niche-based modelling techniques, using the same five bioclimatic variables², to distributional data for a representative sample of European plant diversity (1,350 endemic and non-endemic plant species) under a similar range of climate scenarios to those of Thomas *et al.*¹. To estimate extinction risk, we used each of the three methods employed by them and compared two scenarios of dispersal abilities: universal and none (Table 1). Thomas *et al.*¹ consider differences in extinction predictions between a range of climate-warming scenarios, but our analyses indicate that differences might be at least as strong between models. For example, when using method (3) under a maximum expected warming scenario, predictions from the four models were in the range 2–4.2% with universal dispersal, and in the range 2.3–10.1%

with no dispersal. By contrast, when using method (3) and only one model (generalized linear model), the range for predictions across the three climate scenarios was reduced: a range of 2.7–3.6% with universal dispersal, and 8.2–10.0% with no dispersal.

Second, although Thomas *et al.*¹ show (their Table 4) that their models are highly sensitive to the 'slope' (*z* value) of the species–area relationship, neither their models nor ours yet provide any means of quantifying the uncertainty arising from the simplistic link between proportionate reduction in area and extinction likelihood. Cases of long-term species persistence in remarkably small ranges (for example, on mountain tops and oceanic or land-bridge islands³) demonstrate that, although range reduction is a key driver of species decline, we need to investigate the scale-sensitivity of model outputs and translate projections of range reduction into projections of species losses.

These uncertainties mean that the range of possible extinction risks arising from climate change may be even wider than that reported by Thomas *et al.*¹.

Wilfried Thuiller*†, **Miguel B. Araújo‡**,
Richard G. Pearson‡, **Robert J. Whittaker‡**,
Lluís Brotons*, **Sandra Lavorel§**

*Centre d'Ecologie Fonctionnelle et Evolutive,
CNRS, 34293 Montpellier cedex 5, France

†Climate Change Research Group, Kirstenbosch
Research Center, National Botanical Institute,
Postbag x7, Claremont 7735, Cape Town, South
Africa

e-mail: thuiller@nbi.ac.za

‡Biodiversity Research Group, School of Geography
and the Environment, Oxford University, Oxford
OX1 3TB, UK

§Laboratoire d'Ecologie Alpine, CNRS, Université
Joseph Fourier, BP 53, 38041 Grenoble cedex 9, France
doi:10.1038/nature02716

1. Thomas, C. D. *et al.* *Nature* **427**, 145–148 (2004).

2. Thuiller, W. *Global Change Biol.* **9**, 1353–1362 (2003).

3. Brown, J. H. & Lomolino, M. V. *Biogeography* 2nd edn (Sinauer,
Sunderland, Massachusetts, 1998).

4. Nakicenovic, N. & Swart, R. *Emissions Scenarios: A special report
of working group III of the Intergovernmental Panel on Climate
Change* (Cambridge Univ. Press, Cambridge, 2000).

Reply: Thomas *et al.* reply to this communication
(doi:10.1038/nature02719).

Table 1 Projected percentage extinctions for different models

Models	Universal dispersal			No dispersal		
	Expected climate change					
	Minimum	Mid-range	Maximum	Minimum	Mid-range	Maximum
ANN	2.0, 2.5, 3.0	0, 2.9, 3.7	2.6, 3.3, 3.9	4.8, 6.7, 7.5	6.1, 7.5, 8.7	5.8, 8.1, 9.2
GLM	2.3, 2.4, 2.7	2.5, 2.7, 3.1	3.0, 3.1, 3.6	6.3, 7.5, 8.2	6.8, 8.1, 9.0	7.5, 8.9, 10.0
GAM	2.4, 2.8, 3.2	2.6, 3.2, 3.8	3.1, 3.7, 4.2	5.6, 7.3, 8.4	5.9, 7.8, 9.1	6.7, 8.7, 10.1
CTA	0.9, 2.0, 1.7	1.8, 2.7, 2.7	1.1, 2.3, 2.0	2.0, 5.3, 5.5	2.9, 6.0, 6.5	2.3, 5.9, 6.3

Projected percentage extinction values, based on species–area (for $z = 0.25$) methods¹. The three species–area estimates are ordered in each cell, with method (1) given first, followed by method (2), then method (3). ANN, artificial neural networks; GLM, generalized linear models; GAM, generalized additive models; CTA, classification tree analysis. Experiments for minimum, mid-range and maximum expected climate change were done at the UK Hadley Centre for Climate Prediction and Research for three emission scenarios (A2, B2 and A1, IPCC 2000; ref. 4).

Biodiversity conservation

Effects of changes in climate and land use

Arising from: C. D. Thomas *et al.* *Nature* **427**, 145–148 (2004)

Thomas *et al.*¹ argue, contrary to Sala *et al.*², that climate change poses an equal or greater threat to global biodiversity than land-use change. We contest this claim, however, on the grounds that Thomas *et al.* incorrectly apply species–area relationships.

The species–area relationship ($S = cA^z$, where S is the number of species, A is area, c and z are constants, and z is typically 0.25) predicts the number of species in an independently specified area. Instead, Thomas *et al.*¹ use the sum of the ranges of the species themselves to define the sampling area. Their method is circular because the sum of the ranges of S species is automatically correlated with S itself.

The problem with their entire species–area approach is evident in the first of the several, similar methods that Thomas *et al.*¹ propose. Rearranging the species–area formula shows that extinction risk (fraction of original species projected to become extinct) following reduction of independently specified habitat from A_{original} to A_{new} is $1 - (A_{\text{new}}/A_{\text{original}})^z$. Instead, Thomas *et al.*¹ begin with climate scenarios that predict range loss for individual species. They sum the areas of these species before and after global change to obtain aggregate areas to use in species–area formulae. By arguing that the sum of species' ranges after global change is less than that now, they conclude that extinction risk from global change is large, as great as that from land-use change. This conclusion is premature because the areas used in their calculation are not independent of the number of species. Furthermore, a summation double-counts area occupied by two or more species.

Previous work suggests that species'

ranges have responded individually to historical climatic changes³. Records of historical climatic changes also show that extinction risk is unevenly distributed with respect to range size⁴. Hence, summing ranges does not correctly aggregate extinction risk across species because each species is equally weighted in the extinction-risk calculation. Future estimates of extinction risk might be based on working out the number of species whose areas will drop below a critical patch size because of climate change.

The effects of global change on extinction risk are difficult to anticipate. Global warming will increase some habitats and their species-holding capacity, just as warming reduces other habitats. The net effect for biodiversity of these habitat expansions and contractions is not obvious, particularly as species ranges may shift poleward from the tropics⁵, where the greatest number of species is currently. Although we contest the species–area approach used by Thomas *et al.*¹, we acknowledge species' vulnerability to extinction from climate change⁶. And economic impacts from climate change also exist: for example, the dislocations caused by sea-level rise and the costs of adjusting to new climate cycles.

Lauren B. Buckley, Joan Roughgarden

Department of Biological Sciences, Stanford University, California 94305, USA

e-mail: lbuckley@stanford.edu

doi:10.1038/nature02717

1. Thomas, C. D. *et al.* *Nature* **427**, 145–148 (2004).

2. Sala, O. E. *et al.* *Science* **287**, 1770–1774 (2000).

3. Graham, R. W. *et al.* *Science* **272**, 1601–1606 (1996).

4. Jablonski, D. *Science* **231**, 129–133 (1986).

5. Root, T. L. *et al.* *Nature* **421**, 57–60 (2003).

6. Williams, S. E., Bolitho, E. E. & Fox, S. *Proc. R. Soc. Lond. B* **270**, 1887–1892 (2003).

Reply: Thomas *et al.* reply to this communication (doi:10.1038/nature02719).

Biodiversity conservation

Climate change and extinction risk

Arising from: C. D. Thomas *et al.* *Nature* **427**, 145–148 (2004)

Thomas *et al.*¹ have carried out a useful analysis of the extinction risk from climate warming. Their overall conclusion, that a large fraction of extant species could be driven to extinction by expected climate trends over the next 50 years, is compelling: it adds to the many other reasons why new energy policies are needed to reduce the pace of warming.

There is one important reason, however, why Thomas *et al.* could be greatly underestimating the threat to biodiversity from climate change: genetic adaptation to climate at the population level. Thomas *et al.*¹ assume that if the future climate conditions at any particular location within the current range of a species fall outside the climate envelope determined for the entire species, then that location is no longer a suitable habitat for the species. Implicit is the assumption that all individuals within a species are adapted to the same climate envelope — an envelope determined from the current range of the species.

But now consider a species comprising distinct populations that occupy a subset of ranges arrayed along a climate gradient. If each population can survive only within the relatively narrow climate envelope characterizing its current sub-range, then every individual within the species could suffer stress from climate change². If the individuals cannot migrate, then all individuals within the species would be at risk. By contrast, if existing differential population-level adaptations to climate are ignored, as by Thomas *et al.*, then individuals located towards the cooler end of the current range are unlikely to feel the stress of warming, whether they can migrate or not.

The prevalence of such genetic adaptation to climate at the population level is unknown. Thomas *et al.* could be overestimating extinction rates if evolutionary adaptation to climate change occurs in the future, lessening the effect of warming on

species survival. But the longer generation times of macroscopic species will probably prevent adaptation rates from keeping pace with anthropogenic climate change.

Thomas *et al.* use a species-level approach for estimating extinction under climate change, in contrast to the more traditional community-level approach that uses the species–area relationship (SAR). Although their methods find inspiration from the power-law form of the community-level SAR, instead of enforcing the SAR at the community level before and after climate change, they aggregate information on species-level range areas into an estimate of extinction for the community. This approach is based on the reasonable assumption that species differ in their response to the effects of climate change. However, their modifications of the classical community-level approach for estimating extinction use a common species–area exponent z for all species, which may not be justified. Species that are sparsely distributed within their climate envelope may be less likely to survive a given fractional change in range than species that are found throughout their climate envelope. Deriving predictions for the community by applying z to all species can yield poor results³.

Species differences in the proportional change to their distributional area also arise under land-use change. If the species-level methods of Thomas *et al.* were used to estimate extinctions from land-use change, they would yield different estimates from the traditionally used community-level SAR. As noted by Thomas *et al.*, more work is needed to understand how extinction risk for a given species depends on its reduction in range area.

Estimation methods that involve the application of the SAR (or SAR-inspired species-level equations) to regions with holes or fragmentation typically ignore the confounding effects of shape in the dependence of species richness on area⁴ and, for

this reason, may be inaccurate. An alternative estimation method that could be more accurate in some circumstances demonstrates the degree to which shape influences the expected number of extinctions. This relies on the endemics–area relationship (EAR), which quantifies how the number of species spatially confined to a habitat patch depends on the area of that patch^{3,5,6}.

Where the SAR approach asks how many species are present in the remaining habitat, the EAR approach asks how many species are lost because those species were initially present only in the lost habitat. If shape did not influence either, these methods would yield identical results. However, because it does, EAR can yield significantly different answers from the SAR method, depending on the shape and fragmentation of the lost and remaining habitat^{5,6}. Sorting out where the accurate estimate falls between the EAR and SAR estimates requires as-yet unavailable knowledge of how large-scale geographic patterns of diversity depend on shape and fragmentation.

Thomas *et al.* have focused attention on an important issue. The challenge is to improve our understanding of population genetics and climate tolerance within species, as well as to find the most reliable way to use spatial patterns of diversity and species-level changes in range area for generating estimates of species loss under climate change.

John Harte*, Annette Ostling*, Jessica L. Green†, Ann Kinzig‡

*Energy and Resources Group, University of California, Berkeley, California 94720, USA
e-mail: jharte@socrates.berkeley.edu

†School of Natural Sciences, University of California, Merced, PO Box 2039, California 95344, USA

‡Department of Biology, Arizona State University, PO Box 871501, Tempe, Arizona 85287-1501, USA
doi:10.1038/nature02718

1. Thomas, C. D. *et al.* *Nature* **427**, 145–148 (2003).

2. Jensen, D. Thesis, Univ. California, Berkeley (1993).

3. Green, J. L., Harte, J. & Ostling, A. *Ecol. Lett.* **6**, 919–928 (2003).

4. Kunin, W. E. *Biol. Cons.* **82**, 369–377 (1997).

5. Kinzig, A. & Harte, J. *Ecology* **81**, 3305–3311 (2000).

6. Green, J. L., Harte, J. & Ostling, A. in *Biotic Homogenization* (eds Lockwood, J. L. & McKinney, M. L.) 179–200 (Kluwer Academic, New York, 2001).

Reply: Thomas *et al.* reply to this communication (doi:10.1038/nature02719).

Thomas *et al.* reply — We reconsider our estimates of climate-related extinction¹ in the light of three questions raised by Thuiller *et al.*², Buckley and Roughgarden³ and Harte *et al.*⁴. We are able to confirm our original conclusion that climate change represents a major threat to terrestrial species¹.

First, how much do estimates of extinction depend on the specific approach used to model current and future species distributions? From Table 1 in Thuiller *et al.*², we calculate that distribution model choice produces 1.33-fold variation, on average, in projected extinction in pairwise comparisons of methods. This is important, but smaller than both the twofold variation due to contrasting dispersal assumptions and the twofold variation due to the different climate-change scenarios that we reported¹. The number of environmental variables used during modelling could also affect conclusions, but we found no correlation between these and our estimates of extinction risk in global samples ($r^2 = 0.001$; not significant; each study: 1 d.f.).

Results consistent with our original estimates can be derived from simple population projections without the need for formal distribution models. An index of total population sizes for 12 endemic bird species in Queensland, Australia, has been derived by measuring population density in elevation bands and multiplying by the habitat area in each band (S.E.W. and Luke P. Shoo, manuscript in preparation). Population change was projected directly using the relationship

between elevation and temperature (Australian Meteorological Bureau data). The resulting rate of decline in population size was very similar to the rate of decline in distribution area projected using BIOCLIM (S.E.W. and Luke P. Shoo, manuscript in preparation), which is a distribution modelling method we used to project extinctions¹.

Overall estimates of area loss and extinction risk do not seem to be particularly sensitive to model details when averaging over many species, as we did¹. By contrast, the details can be critical when considering the prognosis for an individual species (for example, when explaining why some species survive in small areas² but others die out). We fully agree that the relative merits of different approaches need further assessment.

Second, can modified species–area relationship methods (SAR) be used to translate average range-area losses into estimates of extinction^{2–4}? We do not accept Buckley and Roughgarden's assertion that our analysis is circular³. As they note, summed area (ΣA) is correlated with the number of species: $\Sigma A = SR$ where S is the number of species and R is the mean range-area of species, either before (R_{original}) or after (R_{new}) some environmental change. We can substitute SR_{new} and SR_{original} in place of ΣA_{new} and $\Sigma A_{\text{original}}$ in our original method (1), such that S cancels out (species with new range sizes of zero are included) and the number of species does not enter into the calculation (for E , the proportion of species projected to become extinct, $E = 1 - [R_{\text{new}}/R_{\text{original}}]^2$). Only mean range change per species matters.

As 50% destruction of a given habitat generates a mean of 50% decline in range area per species, mean area loss per original species (the method that we use) and habitat area loss (as in traditional SAR applications to habitat destruction) on average return the same estimate of expected extinction. None of our three methods is affected by this criticism. Because mean values are used, there is no “double counting” of range areas.

For one vegetation type (the Fynbos floral kingdom in South Africa), we can compare a traditional habitat-based SAR estimate of extinction risk with estimates based on our species-averaging methods. Proteaceae-containing Fynbos vegetation should decline in area by 65% with climate change (climate scenario HadCM2n = Gga[IS92a], for 2050)⁵. The resultant estimate of extinction is 23% (traditional SAR, $z = 0.25$), which is close to our corresponding estimate of 24% of Proteaceae species at risk (mean of three SAR methods, $z = 0.25$, full dispersal, modelled using the same climate scenario and environmental variables)¹. This ‘full dispersal’ comparison is appropriate because traditional SAR applies to habitat area and does not consider additional extinctions caused by range dislocations, when a habitat moves as well as contracts. A third of Fynbos Proteaceae have been predicted to show no overlap between current and future ranges⁵, in line with our estimate of 34% extinction (no dispersal mean of SAR estimates).

Our general conclusions do not even rely on using SAR methods. Harte *et al.*⁴ suggest estimating how many species will lose all of

Table 1 Loss of climatically suitable areas for different species by 2050

Taxon	Region	With dispersal			No dispersal		
		Expected climate change					
		Minimum	Mid-range	Maximum	Minimum	Mid-range	Maximum
Mammals	Mexico ($n = 96$)	0% 0% 10%	0% 4% 15%		3% 6% 39%	3% 10% 37%	
	Queensland ($n = 11$)	0% 9% 36%		64% 73% 100%			
	South Africa ($n = 5$)		20% 40% 80%			40% 60% 100%	
Birds	Mexico ($n = 186$)	0% 0% 4%	0% 1% 8%		0% 0% 11%	1% 1% 8%	
	Europe ($n = 34$)			0% 0% 9%			18% 29% 76%
	Queensland ($n = 13$)	0% 0% 23%		38% 77% 100%			
	South Africa ($n = 5$)		0% 40% 100%			0% 40% 100%	
Frogs	Queensland ($n = 23$)	4% 9% 30%		39% 61% 100%			
Reptiles	Queensland ($n = 18$)	0% 6% 33%		33% 56% 100%			
	South Africa ($n = 26$)		4% 12% 77%			11% 38% 100%	
Butterflies	Mexico ($n = 41$)	0% 2% 17%	0% 5% 34%		0% 2% 7%	2% 2% 12%	
	South Africa ($n = 4$)		0% 0% 25%			50% 75% 100%	
	Australia ($n = 24$)	0% 0% 8%	0% 4% 42%	0% 8% 83%	0% 0% 25%	0% 4% 75%	4% 21% 100%
Other invertebrates	South Africa ($n = 10$)		10% 10% 50%			70% 80% 100%	
Plants	Amazonia ($n = 9$)			78% 78% 78%			100% 100% 100%
	Europe ($n = 192$)	1% 1% 7%	1% 1% 10%	1% 1% 14%	3% 3% 26%	3% 3% 36%	4% 8% 51%
	Cerrado ($n = 163$)				4% 50% 99%	10% 74% 100%	
	South Africa						
	Proteaceae ($n = 243$)		1% 17% 72%			7% 35% 100%	
All species		4% 9% 38%	4% 14% 47%	13% 32% 68%	8% 21% 60%	12% 30% 66%	19% 47% 87%
	($n = 604$)		($n = 832$)	($n = 324$)	($n = 702$)	($n = 995$)	($n = 259$)

Estimated percentages of species projected¹ to lose 100% (first number in each cell), >90% (second number) and >50% (third number) of climatically suitable areas by 2050 (see ref. 1 for climate scenarios). “No dispersal” assumes species survive only in areas of overlap between current and projected future ranges. “With dispersal” assumes species track changing climate distributions perfectly. n , Numbers of species in each category. “All species” values are based on the modelled means (weighted by square-root of number of species in each category, using interpolated values for missing taxa/regions); 100% projected loss implies that surviving populations will be living in climatic zones no longer suitable for their long-term persistence. They are predicted to be “committed to extinction” if 2050 conditions persist indefinitely. Extinctions may be caused by physical aspects of the climate or by biotic interactions that are affected by climate.

their range area. We projected that 5%, 8% and 16% (mean of dispersal scenarios) of the species considered would have lost 100% of their climatically suitable area by 2050, for minimum, mid-range and maximum climate warming, respectively (Table 1). Some 15%, 22% and 40% (mean of dispersal scenarios) of species are projected to have lost more than 90% of their climatically suitable areas by 2050 for minimum, mid-range and maximum warming, respectively (Table 1). A comparable increase in temperature is expected between 2050 and 2100 as between now and 2050 (ref. 6), so distribution changes will continue unabated after 2050. We previously assigned species losing 90% of their climatically suitable area by 2050 a 44% chance of eventual extinction using the method (3) that provided the highest estimate of extinction¹. This was modest, considering that most species projected to lose 90% of their suitable range by 2050 will subsequently lose the remainder as a result of continuing climate change.

Third, how will local adaptations affect the ability of species to respond to climate change, and why will species not be able to evolve adaptations to new conditions, rather than become extinct? Harte *et al.*⁴ are correct

in their assertion that distribution models implicitly assume that locally-adapted regional populations are capable of evolving up to, but not beyond, the set of conditions inhabited by the species as a whole, and they may not always achieve this^{7–9}. Evolving to exploit conditions outside those currently used by the entire species may be difficult because asymmetric gene-flow from distribution cores to margins and/or lack of appropriate variation in marginal populations can prevent the establishment of adaptations that would allow them to colonize ever more extreme environments¹⁰. Why, otherwise, are range sizes for most taxa very small¹¹? In practice, the Quaternary record shows that species have typically responded to past climate changes by shifting range, rather than by evolving *in situ*⁷. Evolutionary responses provide additional uncertainty^{7,12}. If Harte *et al.*⁴ are correct about limitations imposed by ecotypic variation, our estimates of extinction risk will be conservative.

Although further investigation is needed into each of these areas, it is unlikely to result in substantially reduced estimates of extinction. Anthropogenic climate change seems set to generate very large numbers of species-level extinctions.

Chris D. Thomas*, Stephen E. Williams, Alison Cameron, Rhys E. Green, Michel Bakkenes, Linda J. Beaumont, Yvonne C. Collingham, Barend F. N. Erasmus, Martinez Ferreira de Siqueira, Alan Grainger, Lee Hannah, Lesley Hughes, Brian Huntley, Albert S. van Jaarsveld, Guy F. Midgley, Lera Miles, Miguel A. Ortega-Huerta, A. Townsend Peterson, Oliver L. Phillips

**Department of Biology (Area 18), University of York, PO Box 373, York YO10 5YW, UK
e-mail: cdt2@york.ac.uk
doi:10.1038/nature02719*

1. Thomas, C. D. *et al.* *Nature* **427**, 145–148 (2004).
2. Thuiller, W. *et al.* *Nature* doi:10.1038/nature02716 (2004).
3. Buckley, L. B. & Roughgarden, J. *Nature* doi:10.1038/nature02717 (2004).
4. Harte, J., Ostling, A., Green, J. L. & Kinzig, A. *Nature* doi:10.1038/nature02718 (2004).
5. Midgley, G. F., Hannah, L., Millar, D., Rutherford, M. C. & Powrie, L. W. *Global Ecol. Biogeog.* **11**, 445–451 (2002).
6. Houghton, J. T. *et al.* *Climate Change 2001: The Scientific Basis. Contributions of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, 2001).
7. Davis, M. B. & Shaw, R. G. *Science* **292**, 673–679 (2001).
8. Pimm, S. L. *Nature* **411**, 531–532 (2001).
9. Thomas, C. D. *et al.* *Nature* **411**, 577–581 (2001).
10. Kirkpatrick, M. & Barton, N. H. *Am. Nat.* **150**, 1–23 (1997).
11. Gaston, K. J. *The Structure and Dynamics of Geographic Ranges* (Oxford Univ. Press, Oxford, 2003).
12. Holt, R. D. *Evol. Ecol. Res.* **5**, 159–178 (2003).

Genome evolution in yeasts

Bernard Dujon¹, David Sherman^{5,6}, Gilles Fischer¹, Pascal Durrens^{6,7}, Serge Casaregola⁸, Ingrid Lafontaine¹, Jacky de Montigny⁹, Christian Marck¹⁰, Cécile Neuveglise⁸, Emmanuel Talla¹, Nicolas Goffard⁶, Lionel Frangeul², Michel Aigle⁷, Véronique Anthouard¹¹, Anna Babour⁸, Valérie Barbe¹¹, Stéphanie Barnay⁸, Sylvie Blanchin⁸, Jean-Marie Beckerich⁸, Emmanuelle Beyne^{5,6}, Claudine Bleykasten⁹, Anita Boisramé⁸, Jeanne Boyer¹, Laurence Cattolico¹¹, Fabrice Confanioli¹², Antoine de Daruvar⁶, Laurence Despons⁹, Emmanuelle Fabre¹, Cécile Fairhead¹, Hélène Ferry-Dumazet⁶, Alexis Groppi⁶, Florence Hantraye³, Christophe Hennequin¹, Nicolas Jauniaux⁹, Philippe Joyet⁸, Rym Kachouri¹³, Alix Kerrest¹, Romain Koszul¹, Marc Lemaire¹⁴, Isabelle Lesur⁵, Laurence Ma², Héloïse Muller¹, Jean-Marc Nicaud⁸, Macha Nikolski⁵, Sophie Oztas¹¹, Odile Ozier-Kalogeropoulos¹, Stefan Pellenz¹, Serge Potier⁹, Guy-Franck Richard¹, Marie-Laure Straub⁹, Audrey Suleau⁸, Dominique Swennen⁸, Fredj Tekai¹, Micheline Wésolowski-Louvel¹⁴, Eric Westhof¹³, Bénédicte Wirth⁹, Maria Zeniou-Meyer⁹, Ivan Zivanovic¹², Monique Bolotin-Fukuhara¹², Agnès Thierry¹, Christiane Bouchier², Bernard Caudron⁴, Claude Scarpelli¹¹, Claude Gaillardin⁸, Jean Weissenbach¹¹, Patrick Wincker¹¹ & Jean-Luc Souciet⁹

¹Unité de Génétique Moléculaire des Levures (URA 2171 CNRS and UFR 927 Université Pierre et Marie Curie), ²Plate-forme génomique, Pasteur Génomole Ile-de-France, ³Unité de Génétique des interactions macromoléculaires (URA 2171 CNRS), and ⁴Groupe Logiciels et Banques de données, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France

⁵Laboratoire Bordelais de Recherche en Informatique (LaBRI, UMR 5800 CNRS), 351 cours de la Libération, 33405 Talence Cedex, France

⁶Centre de Bioinformatique de Bordeaux, Université Victor Ségalen (Bordeaux 2), 146 rue Léo Saignat, 33076 Bordeaux Cedex, France

⁷Institut de Biochimie et Génétique Cellulaires (UMR 5095 CNRS), Université Victor Ségalen (Bordeaux 2), 1 rue Camille Saint-Saëns, 33077 Bordeaux Cedex, France

⁸Collection de Levures d'Intérêt Biotechnologique et Laboratoire de Génétique Moléculaire et Cellulaire (UMR 216 INRA and URA 1925 CNRS), INA-PG, PO Box 01, 78850 Thiverval-Grignon, France

⁹Laboratoire de Dynamique, Evolution et Expression des Génomes de Microorganismes (FRE 2326 CNRS), Université Louis Pasteur, 28 rue Goethe, 67000 Strasbourg, France

¹⁰Service de Biochimie et de Génétique Moléculaire, CEA/Saclay, 91191 Gif-sur Yvette, France

¹¹Génoscope (UMR 8030 CNRS), 2 rue Gaston Crémieux, 91057 Evry Cedex, France

¹²Institut de Génétique Moléculaire (UMR 8621 CNRS), Université de Paris Sud, Bâtiment 400, 91405 Orsay Cedex, France

¹³Modélisations et Simulations des Acides Nucléiques, IBMC (UPR 9002 CNRS), 15 rue René Descartes, 67000 Strasbourg, France

¹⁴Laboratoire de Génétique des Levures (UMR 5122 CNRS), Université Claude Bernard, Bâtiment Lwoff, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Identifying the mechanisms of eukaryotic genome evolution by comparative genomics is often complicated by the multiplicity of events that have taken place throughout the history of individual lineages, leaving only distorted and superimposed traces in the genome of each living organism. The hemiascomycete yeasts, with their compact genomes, similar lifestyle and distinct sexual and physiological properties, provide a unique opportunity to explore such mechanisms. We present here the complete, assembled genome sequences of four yeast species, selected to represent a broad evolutionary range within a single eukaryotic phylum, that after analysis proved to be molecularly as diverse as the entire phylum of chordates. A total of approximately 24,200 novel genes were identified, the translation products of which were classified together with *Saccharomyces cerevisiae* proteins into about 4,700 families, forming the basis for interspecific comparisons. Analysis of chromosome maps and genome redundancies reveal that the different yeast lineages have evolved through a marked interplay between several distinct molecular mechanisms, including tandem gene repeat formation, segmental duplication, a massive genome duplication and extensive gene loss.

Comparative genomics in eukaryotes has been limited by the considerable phylogenetic distances between the first sequenced organisms: a yeast (*S. cerevisiae*)¹, a nematode (*Caenorhabditis elegans*)², an insect (*Drosophila melanogaster*)³, a dicotyledonous plant (*Arabidopsis thaliana*)⁴ and man (*Homo sapiens*)⁵. New sequencing programmes have recently extended these comparisons to organisms of the same phylogenetic group, but the number of completely sequenced genomes remains low, and intriguing differences appear between the groups. For example, two sequenced Diptera (*Anopheles gambiae*⁶ and *D. melanogaster*) believed to have diverged from their common ancestor about 250 million years (Myr) ago show a larger sequence divergence than two vertebrate species (*H. sapiens* and the fish *Takifugu rubripes*⁷), which diverged 450 Myr ago. The recently published draft sequence of the *Caenorhabditis briggsae* genome reveals extensive map collinearity with *C. elegans*, from which it diverged 100 Myr⁸ ago. By contrast, two cultivars of rice, *Oryza sativa*^{9,10}, show a very limited synteny with *A. thaliana*, from which they diverged 200 Myr ago.

Yeasts and fungi are ideal organisms for comparative genomic studies in eukaryotes because of their small and compact genomes and because they include a number of species, such as *Neurospora crassa*¹¹, *S. cerevisiae* and *Schizosaccharomyces pombe*¹², that have

been, and continue to be, used extensively in genetic studies. However, the divergence between these three species is ancient (estimated to be at least 300 Myr old) and the organization of their genomes is quite different. The diversity of the hemiascomycetes—a group of ascomycetes that contains most of the known yeast species—was first explored four years ago using low-coverage sequencing of 13 distinct species¹³. More recently, deeper sequencing coverages were applied to a few *Saccharomyces* species very closely related to *S. cerevisiae* (the *sensu stricto* group), plus one more distant species, *Saccharomyces kluyveri*, in order to identify conserved regulatory elements^{14,15}. While this article was submitted, the complete genome sequences of *Ashbya gossypii*¹⁶, a filamentous yeast, and *Kluyveromyces waltii*¹⁷ were used to map and analyse the ancient genome duplication in the ancestry of *S. cerevisiae*.

Instead of focusing on closely related species or on the origin of *S. cerevisiae*, we decided to explore the evolution of the hemiascomycete phylum as broadly as possible. On the basis of our previous estimates¹³ we selected four yeast species representing various and distant branches among the hemiascomycetes for complete sequencing. *Candida glabrata* was chosen because it has become the second causative agent of human candidiasis, and because, despite its name, it is phylogenetically more closely related to *S. cerevisiae*

than to *C. albicans*, the major human fungal pathogen with which it shares only a few properties. *Kluyveromyces lactis* is a yeast species commonly used for genetic studies, and it occupies an interesting position within the phylogeny of hemiascomycetes. *Debaryomyces hansenii* was selected because it is a halotolerant yeast, related to *C. albicans* and other pathogenic yeasts, that is often found on fish and salted dairy products. *Yarrowia lipolytica*, an alkane-using yeast commonly used in genetic studies, is very distantly related to the rest of the yeasts; instead it shares a number of common properties with filamentous fungi. For each species, the haploid type strain was sequenced. Of importance for evolutionary studies, the four yeast species display different mechanisms of sexuality (see ref. 13). *Yarrowia lipolytica* has a haplo-diplontic cycle (that is, it alternates between haploid and diploid phases of similar importance), whereas *D. hansenii* is a homothallic yeast with an essentially haplontic life cycle. Both species have only one mating-type locus (*MAT*), whereas the other two have two silent mating-type cassette homologues, similar to *S. cerevisiae*. As is often the case with pathogens, *C. glabrata* displays no known sexual cycle, despite the fact that haploid strains of the two distinct mating types are regularly isolated from patients. Finally, *K. lactis* is a heterothallic species with a predominantly haplontic cycle, in contrast to *S. cerevisiae* in which the predominantly diplontic cycle is pseudo-heterothallic owing to mating-type switching.

This work, which represents the first multispecies exploration of genome evolution across an entire eukaryotic phylum, reveals the variety of events and mechanisms that have taken place, and should allow useful comparisons with other phyla of multicellular organisms when more genome sequences are determined.

Overview of the four yeast genomes

Sequencing status of the four yeast species is summarized in Table 1. Genome sizes and chromosome numbers vary within an approximately twofold range between the four species, but without direct correlation. In contrast, the total number of protein-coding genes varies only by 1.3 times between the four species (Table 2), whereas the total number of transfer RNA genes varies by more than three times between *Y. lipolytica* (510 genes) and *K. lactis* (162 genes). The overall gene density is significantly lower in *Y. lipolytica* (one gene per 3 kilobases (kb)) than in the other yeasts (one gene per 2 kb, as in *S. cerevisiae*). We have identified all of the centromeres from

C. glabrata, *K. lactis* and *Y. lipolytica* (see Supplementary Table S1), but were unable to detect the centromeres of *D. hansenii*. We also identified telomeric repeats and proteins of the telomerase complex (see Supplementary Table S2). Transposable elements will be reported elsewhere (S.C., C.N. and P.W., unpublished data).

Non-coding RNA genes

Ribosomal RNA genes

There is a large diversity in the organization of rDNA repeats among yeasts. Compared with the single intrachromosomal rDNA repeat locus of *S. cerevisiae* and *K. lactis*, three distinct intrachromosomal loci are found in *D. hansenii*, whereas seven and two loci are found in subtelomeric regions of *Y. lipolytica* and *C. glabrata*, respectively. Variability also prevails for the 5S rRNA gene copies. In contrast with *S. cerevisiae* and *K. lactis*, where a single copy of this gene is located in opposing orientation between each repeat unit of the 35S primary transcript (the precursor of the 18S, 5.8S and 26S rRNA molecules), two copies occur in tandem in each repeat in *C. glabrata* and *D. hansenii*, and the same gene is dispersed in 105 copies (plus 11 pseudogenes) throughout the genome of *Y. lipolytica*.

Transfer RNA genes

The type and number of tRNA genes (tDNA) are described in Supplementary Table S3. *Candida glabrata* and *K. lactis* display exactly the same 42-tDNA set as *S. cerevisiae*, whereas *D. hansenii* uses a slightly different 43-tDNA set, and *Y. lipolytica* uses the same 44-tDNA set as higher eukaryotes¹⁸. The CUG codon (leucine) is used as a serine codon in *D. hansenii*¹³, and is read by the special single-copy tRNA-Ser (CAG), as in *C. albicans*. Note that this modification of the genetic code does not exist in *C. glabrata*, another indication of the artificial nature of this genus. Introns are found in about one-quarter of the tDNAs in *C. glabrata*, *K. lactis* and *D. hansenii* (see Supplementary Table S4), as in *S. cerevisiae*; however, in *Y. lipolytica* 26 of the 44 tDNAs contain introns, a proportion that has not been observed so far in other eukaryotes¹⁸. As in *S. cerevisiae*, tRNA genes are scattered throughout the genomes of the four yeast species. We found no gene clusters, except in *D. hansenii* where eight identical copies of a tDNA-Lys (CTT) are repeated in tandem separated by intergenic distances sufficient for independent transcription (188–1,855 bp). Notably, a number of tDNA pairs are observed in which the distance separating the two

Table 1 Genome assemblies of the four yeast species

Species	Strain	Number of chromosomes	Total reads	Coverage (sequence)	Coverage (clones)	N50 contigs (kb)	N50 scaffolds (kb)	Total gaps	Assembly size (without rDNA) (kb)
<i>C. glabrata</i>	CBS138	13	188,853	× 8	× 30	1,000	1,025	6	12,280
<i>K. lactis</i>	CLIB210	6	152,071	× 11.4	× 56	1,670	1,670	0	10,631
<i>D. hansenii</i>	CBS767	7	150,570	× 9.7	× 36	102	2,038	207	12,221
<i>Y. lipolytica</i>	CLIB99	6	247,279	× 10	× 59	704	3,453	11	20,503

Sequencing and assembly were performed as described in Methods. Sequences of *C. glabrata*, *K. lactis* and *Y. lipolytica* are finished (no gap) or contain very few gaps. The sequence of *D. hansenii* is in the form of a high-quality draft. In all cases, each chromosome of each yeast is either complete (single contig) or represented by a single super-contig (scaffold). Most remaining gaps are small or contain repeated sequences. Some subtelomeric regions are missing from the assembly because they are too similar to one another to be assigned to a specific chromosome. rDNA repeats are assembled separately. N50, median values.

Table 2 General characteristics of the yeast genomes and predicted proteomes

Species	Genome size (Mb)	Average G+C content (%)	Total CDS	Total tRNA genes	Average gene density (%)	Average G+C in CDS (%)	Average CDS size (codons)	Median CDS size (codons)	Maximum CDS size (codons)
<i>S. cerevisiae</i>	12.1	38.3	5,807	274	70.3	39.6	485	398	4,911
<i>C. glabrata</i>	12.3	38.8	5,283	207	65.0	41.0	493	409	4,881
<i>K. lactis</i>	10.6	38.7	5,329	162	71.6	40.1	461	381	4,916
<i>D. hansenii</i>	12.2	36.3	6,906	205	79.2	37.5	389	307	4,190
<i>Y. lipolytica</i>	20.5	49.0	6,703	510	46.3	52.9	476	399	6,539

Figures are calculated from final chromosome sequences or scaffolds, after annotation. Genome sizes do not include rDNA. Average gene density represents the fraction of each genome occupied by the protein-coding genes (other genetic elements are not considered). Figures for *D. hansenii* are only tentative; figures for *S. cerevisiae* were recently recomputed from <http://mips.gsf.de/genre/proj/yeast>.

Table 3 Classification of yeast proteins in families

Family class	Number of families	Number of CDS					
		SACE	CAGL	KLLA	DEHA	YALI	Total
Robust (identical + reconciled)	3,410	4,094	3,651	3,504	3,832	3,296	18,377
Consensus	1,311	1,287	1,201	1,176	1,831	1,894	7,389
Non-assigned	—	426	431	649	1,243	1,513	4,262
Total	4,721	5,807	5,283	5,329	6,906	6,703	30,028

The table shows the total number of protein families in each class and the corresponding numbers of CDS in each yeast species. Families were classified as explained in Methods. See Fig. 1 for species abbreviations.

genes is shorter than the minimal 5' sequence required for transcription (see Supplementary Table S5). Consistent with the idea of co-transcription, the two members of each pair are always co-oriented. Such structures must result from independent formation followed by subsequent duplications in each phylogenetic branch, as judged from the fact that they are distinct for each species and are often present in multiple copies dispersed throughout the genome.

Other non-coding RNA genes

Most RNA-polymerase-III-transcribed genes and other non-coding RNA genes are not well conserved in yeasts. Nevertheless, we have identified the U1–U6 small nuclear RNAs, as well as the RNA components of the RNase P, the signal recognition particle (SRP) and, in two species, the telomerase complex (see Supplementary Table S6). Some RNAs show extensive size variation between species. Although most RNA genes are unique in each yeast, the U3 gene is duplicated in *D. hansenii* and is triplicated in *Y. lipolytica*; the U1 and SRP genes are duplicated in *Y. lipolytica*; and the U4 gene is duplicated in tandem in *D. hansenii*.

Protein families and genome redundancy

Classification of yeast proteins and sequence conservation

Together with previous data from *S. cerevisiae*¹, our four new yeast sequences offer a unique collection of 30,028 proteins from five phylogenetically related species. As a first step towards identifying homologies between these genomes, the proteins were classified into families of homologues (see Methods and Table 3). When sorted according to their phyletic patterns (presence or absence in each species, Table 4), over 40% of the families (2,014 families and 17,153 genes) are common to the five yeasts (pattern 'sckdy', where 's' indicates *S. cerevisiae*, 'c' indicates *C. glabrata*, and so on). Most of these 'universal' families (1,208) contain a single protein from each species (1:1:1:1:1 relationship), considered here as 'direct orthologues'. The remaining 806 universal families contain at least one paralogous pair in at least one species, creating various situations of paralogy subtypes as defined in ref. 19. In agreement with the phylogeny of the five yeast species (see Supplementary Fig. S7), the next most abundant patterns are sck– and –dy, followed by sckd–. All families common to our five yeasts (pattern sckdy) have homologues in other hemiascomycete species, 98% of them have homologues in other ascomycetes, and 92% have homologues in basidiomycetes (see Table 4). Thus, most of these proteins seem to be universally or largely conserved in evolution. Evolutionary conservation is not as widespread for non-sckdy protein families, but remains consistent with the phylogeny (for example, compare protein families in the sck– pattern, which are poorly conserved beyond yeasts, with those in the –dy pattern, which are more often conserved beyond yeasts). In total, approximately 800 homologous protein families seem to be ascomycete-specific, of which about 660 are specific to hemiascomycetes only (not shown). The functions of these families seem to be highly diversified.

Overall genome redundancy as deduced from protein families

The global degree of genome redundancy, estimated from the number of paralogous gene copies per protein family, illustrates the importance of ancestral gene duplications in all yeasts (Fig. 1). Quantitative variations are, however, observed between species. *Kluyveromyces lactis* has the least duplicated genome (501 sets of

Table 4 Phyletic patterns of yeast protein families and conservation in other fungi

Pattern	Families	Proteins	Conservation (%)			
			Hemiascomycetes	Ascomycetes	Basidiomycetes	All
Families universal to all						
sckdy	2,014	17,153	100	98	92	100
Families restricted to <i>S. cerevisiae</i> , <i>C. glabrata</i> and/or <i>K. lactis</i>						
sck--	572	1,827	100	36	19	100
sc---	245	547	100	19	6	100
s-k--	102	212	100	25	11	99
-ck--	34	77	97	31	19	94
Families restricted to <i>D. hansenii</i> and <i>Y. lipolytica</i>						
---dy	488	1,121	62	77	52	89
Families missing in one species						
-ckdy	17	100	100	94	82	94
s-kdy	49	312	100	98	90	98
sc-dy	44	215	100	93	80	98
sck-y	81	382	100	92	76	99
sckd-	316	1,445	100	79	57	100
Species-specific families						
s----	31	95	90	13	0	87
-c---	8	51	63	50	50	50
--k--	16	38	50	13	6	44
---d-	135	501	42	40	29	55
----y	177	492	31	44	32	45
All other combinations						
sc-d-	16	53	100	68	55	95
sc--y	7	24	100	88	75	88
s-kd-	32	125	100	79	62	97
s--dy	14	46	100	93	87	93
s-k-y	8	27	100	71	57	86
-c-dy	19	63	100	75	50	95
-ckd-	13	43	100	50	33	92
-ck-y	3	9	100	100	100	67
--kdy	103	380	95	86	59	96
s--d-	51	110	98	59	44	96
s--y	7	15	83	83	50	83
-c-d-	18	53	100	77	64	95
-c--y	17	45	82	71	47	82
--kd-	58	143	85	51	33	92
--k-y	26	62	92	88	65	92
Total	4,721	25,766				

The phyletic pattern of each protein family indicates the presence (s, c, k, d, y) or absence (-) of its members in each yeast species (see text for definition of single-letter abbreviations). The total number of families and proteins are indicated. Note that the number of species-specific families may be overestimated because, in the present stage of annotation, some predicted ORFs might not correspond to actual genes. For each protein family a position-specific scoring matrix was computed for two rounds using BLASTppg, with the longest member of the family serving as representative. Each position-specific scoring matrix and protein family representative was used to search a combined data bank of completely or partially sequenced fungal species, using psitBLASTn, with an expected cutoff value of 1×10^{-10} and no filtering of low-complexity regions. A total of 239,464 matches were recorded. The sequenced organisms included in the comparisons are: (1) hemiascomycetes: *Saccharomyces bayanus*, *S. castellii*, *S. kluyveri*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus* and *C. albicans*; (2) other ascomycetes: *Aspergillus fumigatus*, *A. nidulans*, *Coccidioides immitis*, *Fusarium graminearum*, *Magnaporthe grisea*, *N. crassa*, *Pneumocystis carinii*, *P. carinii carinii* and *S. pombe*; and (3) basidiomycetes: *Phanerochaete chrysosporium* and *Ustilago maydis*. The table shows the percentage of families of each pattern conserved in at least one other species of hemiascomycetes, other ascomycetes, basidiomycetes, or all simultaneously (All).

paralogues), whereas *D. hansenii* has the most duplicated one (901 sets). *Saccharomyces cerevisiae* and *Y. lipolytica* have an intermediate and similar level of duplication, and *C. glabrata* stands between *S. cerevisiae* and *K. lactis*. The maximum range of variation in global genome redundancy is only 1.6-fold (31.8% for *K. lactis* versus 51.5% for *D. hansenii*), a figure that, compared to duplication events discussed below, illustrates the importance of ancestral duplications that occurred before divergence of hemiascomycetes and the extent of gene loss in each branch.

Divergence between orthologous and paralogous proteins

The distributions of amino acid sequence divergence between orthologous proteins were calculated from all pairwise comparisons between the five yeast species. All distributions are unimodal, and their mean and median values are consistent with phylogeny (Fig. 2a–c). Notably, despite their similar morphologies and life-styles, the five yeast species appear to be more diverse at the molecular level than, for example, the entire phylum of chordates. The average sequence identity between orthologous proteins of mammals (man or mouse) and fishes (*Takifugu* or *Tetraodon*) is about 70% (refs 7, 20, and O. Jaillon and H. Roest-Crolius, personal communication) compared with only 65% between *S. cerevisiae* and *C. glabrata*, or 60–61% between *K. lactis* and either *S. cerevisiae* or *C. glabrata* (Fig. 2a). Similarly, the average level of amino acid identity between *Y. lipolytica* and any of the other yeasts is 48–49%, a figure similar to that found between the urochordate *Ciona intestinalis* and any of the vertebrates (ref. 21, and O. Jaillon and H. Roest-Crolius, personal communication). Such a broad evolutionary range within the hemiascomycete yeasts was anticipated from our first low-coverage sequence exploration¹³, and justifies *a posteriori* the present selection of species. In contrast to orthologous proteins, sequence divergence between paralogues in each species shows bimodal distributions with an abundance of low sequence identities (25–50%)—probably representing ancient duplications—and a moderate excess of highly similar proteins (90–100%) relative to the intermediate range (Fig. 2d). This last fraction, which is nearly absent in *C. glabrata*, must reflect recent duplications and/or sequence homogenization by gene conversion. Interestingly, the major peaks have similar modes for all five species (roughly 35% identity), indicating that these paralogues largely correspond to duplications having occurred before species divergence. By contrast, the paralogous pairs showing over ~60% identity must largely reflect duplication events that occurred after the speciation of these five yeasts, as this fraction quantitatively varies between species.

Expansion and contraction within universal protein families

Specific gene duplications in some phylogenetic branches and/or gene losses in others are represented by families containing unequal numbers of proteins between species. Such a situation is found in over 700 universal protein families (sckdy). In some cases, adaptive evolution, as reported in *S. cerevisiae*²², or functionally significant gene dosage effects may exist. But as most of these variations concern small numbers of genes (1–3), they cannot be statistically distinguished from random accidents with no specific phenotypic effect. By contrast, 14 protein families were identified showing statistically significant (with a probability of 10^{-3}) expansion in *Y. lipolytica*, *D. hansenii* or both. Judging from their *S. cerevisiae* members, these families encode acylglycerol lipases, proteins similar to sphingomyelinases, α -1,4-glucan glucosidases, alkaline extracellular proteases, GPI-anchored aspartyl proteases, choline or allantoin transporters, peroxisomal 2,4-dienoyl-CoA reductase, C-22 sterol desaturase and other cytochrome P450 enzymes, or NADPH dehydrogenase. One family remains of unknown function. Multigenic families encoding multidrug resistance proteins and hexose transporters are specifically more expanded in *D. hansenii* than in the other four yeasts. In nearly all cases, the family expansions correspond to additional gene copies dispersed in

each genome, suggesting ancient and multiple gene duplication events. Tandem gene repeat formation has a negligible role in family size expansions, except in *D. hansenii* (see below).

Genetic map organization and genome redundancy

Analysis of the genetic maps of the four yeasts and comparisons between them and *S. cerevisiae* reveals an interplay of several distinct mechanisms that have contributed unequally to the evolution of these genomes in the different lineages.

Tandem gene duplications

As in *S. cerevisiae*, a few dozen tandem gene arrays, mostly composed of two or three gene copies, are dispersed throughout the genomes of *C. glabrata*, *K. lactis* and *Y. lipolytica*. However, the genome of *D. hansenii* was found to be much richer in such structures than the other yeast species (Fig. 3). A total of 247 arrays (329 gene pairs) are distributed all over this genome, significantly contributing to its global redundancy. Furthermore, arrays of 8 to 9 repeats are encountered in this species (Fig. 3). The quantitative difference between *D. hansenii* and the other yeasts was unexpected. It may result from a more efficient creation of tandem arrays (or a recent massive episode) or a less efficient destruction of newly formed tandem repeats (for example by ‘pop-out’). Although their sequence divergence is generally lower than that of dispersed paralogues, tandem paralogues are generally not strictly identical in sequence, indicating possible functional specialization (and limiting their destruction by pop-out). Frameshifted pseudogenes and gene fragments are often found in large arrays. Finally, in 27 cases the repeats are made of two genes forming alternating arrays. The functions of genes included in tandem arrays appear to be extremely diverse (see Supplementary Table S8). As only two arrays are common to all yeasts, it is probable that new arrays are constantly formed (and possibly rapidly resolved) in all lineages.

Blocks of ancestral duplications

A significant part of the genome redundancy in *S. cerevisiae* is due to

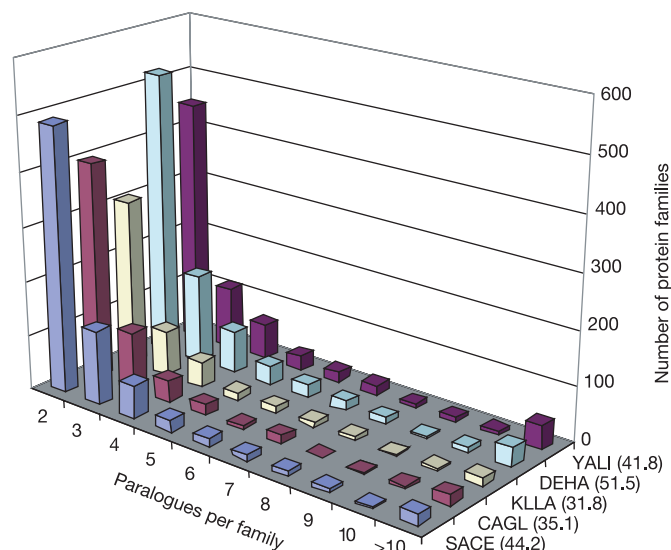


Figure 1 Overall genome redundancy as deduced from protein families. Shown for each yeast species is the total number of protein families distributed according to their size. A similar pattern (not shown) is obtained when considering only the universal protein families (sckdy pattern). The overall genome redundancy for each species, defined as the ratio (in per cent) of the number of CDS belonging to multigene families over the total number of CDS, is indicated in brackets next to the species abbreviation. SACE, *S. cerevisiae*; CAGL, *C. glabrata*; KLLA, *K. lactis*; DEHA, *D. hansenii*; YALI, *Y. lipolytica*.

the presence of sister chromosomal regions or blocks²³ thought to be the remnants of an ancestral whole-genome duplication that occurred after its divergence from *K. lactis* and before its divergence from *C. glabrata*^{24,25}. According to this theory, similar numbers of blocks are expected in *C. glabrata* and *S. cerevisiae*, and no block is expected in other yeasts. Rather than this simple all-to-none transition, our results instead show gradual quantitative differences between the species (Fig. 4). Compared with the 56 blocks scattered throughout the genome map in *S. cerevisiae* (plus 21 blocks in subtelomeric regions), only 20 blocks are found in *C. glabrata* (approximately three times less), 8 blocks (1 tandem) in *K. lactis*, 5 blocks (3 in tandem) in *D. hansenii* and 2 blocks in *Y. lipolytica*. A closer examination of the results reveals that 18 of the 20 blocks identified in *C. glabrata* correspond to blocks also present in *S. cerevisiae* (see Supplementary Table S9). This high coincidence strongly suggests that the duplicated blocks of *C. glabrata* have the same ancestral origin as those of *S. cerevisiae*, indicating that a major duplication occurred before the divergence of the two species and after their divergence from *K. lactis*. Consistently, none of the eight duplicated blocks in *K. lactis* coincide with the above, indicating a distinct formation; the same applies for the few duplicated blocks of *D. hansenii*.

The marked difference between *C. glabrata* and *S. cerevisiae* needs, however, to be explained in order to be reconciled with a whole-genome duplication event in their common ancestry (this

whole-genome duplication is also documented by recent comparisons to *A. gossypii*¹⁶ and *K. waltii*¹⁷). Because the *C. glabrata* blocks tend to be smaller, on average, than those of *S. cerevisiae* (Fig. 4), a plausible explanation for their reduced number is that, after the duplication, a higher rate of gene loss occurred in *C. glabrata* compared with *S. cerevisiae*, thus effacing many blocks inherited from their common ancestor and reducing the size of others. This possibility is consistent with the reductive genome evolution mentioned below for this pathogenic yeast, and with its reduced conservation of synteny with other yeasts (not shown). It is also consistent with the lower global genome redundancy of *C. glabrata* compared with *S. cerevisiae* (see Fig. 1) and with the reduced number of highly similar paralogues (see Fig. 2).

The duplicated blocks observed in *K. lactis*, *D. hansenii* and *Y. lipolytica* are too few to indicate other possible massive genome duplications in their own lineages. Instead, like some of the blocks that do not coincide between *C. glabrata* and *S. cerevisiae*, they may have originated from independent segmental duplications, a mechanism now directly demonstrated experimentally in *S. cerevisiae*²⁶. In general, the absence of a direct correlation between the total number of duplicated blocks in the genome maps and the global levels of genome redundancy judged from individual genes (see above), illustrates the multiplicity of events that have occurred in the various lineages. If an ancestral genome duplication event accounts for a large part of the redundancy of the

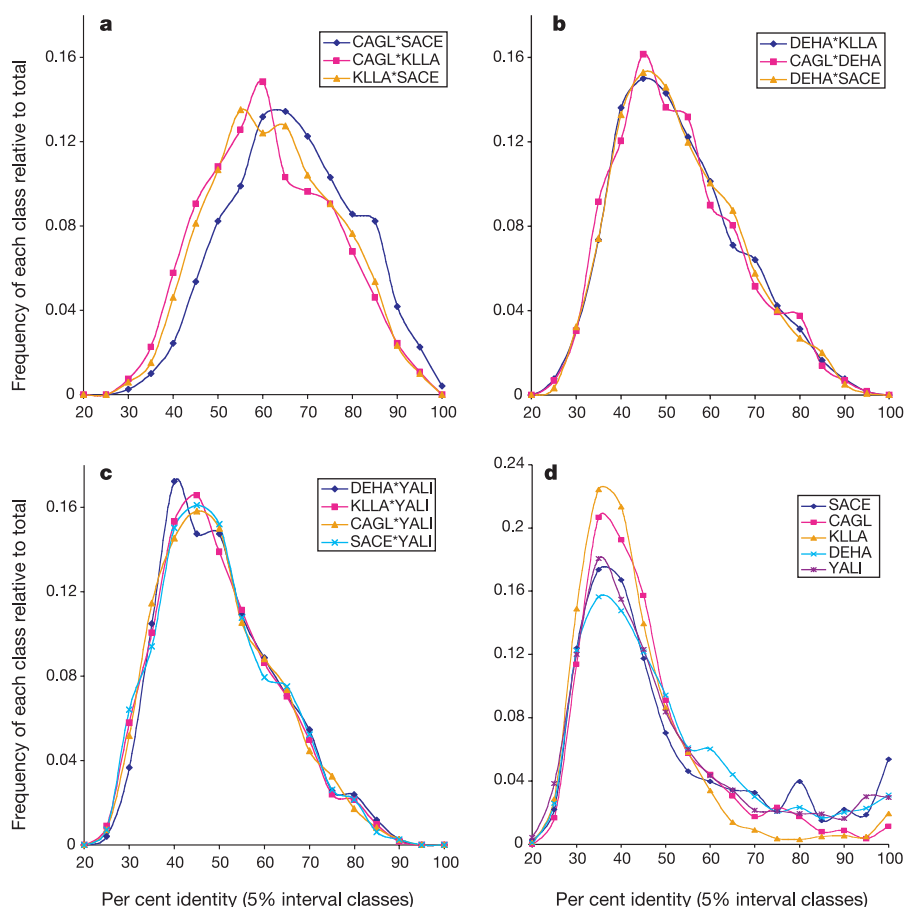


Figure 2 Distribution of the percentage identity between pairs of homologous proteins. Pairs of direct orthologues (a–c) and paralogues (d) defined after classification of the proteins (see text) were used to compute the distributions. Amino acid identities were calculated from Smith–Waterman alignments (see Methods). Each distribution was computed from over 1,100 pairwise alignments (orthologues) and from 2,200 to 5,700

pairwise alignments (paralogues). Mean identity values between all orthologous proteins allow meaningful comparisons with other classes of organisms as indicated in the text (48–49% for all comparisons including *Y. lipolytica* (c), about 51% for the comparisons between *D. hansenii* with the three other yeasts (b), and 60–65% for *S. cerevisiae*, *C. glabrata* and *K. lactis* (a)). Species abbreviations as in Fig. 1.

genomes of *S. cerevisiae* and *C. glabrata*, other yeasts have acquired equivalent or even larger redundancies by other mechanisms (see Discussion).

Conservation of synteny

Studying the conservation of synteny between species is another route to trace the evolutionary events that affect genomes. When applied to yeasts, this exercise reveals the considerable extent of genome reorganization that has occurred in this phylum (Fig. 5). The largest syntenic clusters between any two of our five yeasts contain only about 40 gene pairs. *Saccharomyces cerevisiae*, *C. glabrata* and *K. lactis* share about 500 syntenic clusters between them (considered in pairwise combinations), but the numbers and sizes of clusters rapidly decrease over increasing phylogenetic distances. We found only 74–149 syntenic clusters between *D. hansenii* and any other yeast species, and only 34–90 clusters between *Y. lipolytica* and any of the others. This extensive level of genomic map reshuffling contrasts with the near complete collinearity between genomes of the *Saccharomyces sensu stricto* group^{14,15,27}, but is coherent with the large evolutionary distance between our yeasts deduced from the degree of amino acid replacement between orthologous proteins (see above).

Despite the extent of map rearrangements, comparisons between *K. lactis* and *S. cerevisiae* reveal that 78% of the genes belonging to syntenic clusters correspond to intermingled series in which one region of the *K. lactis* genome corresponds to two (and sometimes more) distinct chromosomal regions in *S. cerevisiae*. Similar results have been reported separately by comparing the genomes of *K. waltii*¹⁷ and *A. gossypii*¹⁶ with *S. cerevisiae*. When *C. glabrata* is

compared to *K. lactis*, only 64% of the genes belonging to syntenic clusters correspond to intermingled series in which one region of the *K. lactis* genome corresponds to two distinct chromosomal regions of *C. glabrata*. Again, this quantitative difference (64% compared with 78%) can only be reconciled with a general duplication event in the common ancestor of *S. cerevisiae* and *C. glabrata* if one assumes a more extensive rate of gene loss in the *C. glabrata* lineage than in *S. cerevisiae*.

Gene dynamics and sequence conservation

In addition to duplication and divergence, the gain and loss of specific genes may be critical for the functional differentiation between species and for evolution. Considering only the families of well conserved proteins between yeasts (to eliminate artificial threshold effects through a gradual increase in sequence divergence when phylogenetic distances increase), we have identified species-specific gene losses using, for criterion, the absence of a protein in one species and its presence in the four others (see Supplementary Table S10). The most striking example of species-specific gene losses is offered by *C. glabrata* where 29 genes are lost compared to all other yeasts. Losses seem to affect genes in a functionally coordinated manner, suggesting a reductive evolutionary scheme, possibly associated with the emergence of this yeast as a human pathogen. Specific losses in *C. glabrata* include genes involved in (1) galactose metabolism (five genes); (2) phosphate metabolism (four genes); (3) cell rescue, defence and virulence (three genes); and (4) nitrogen and sulphur metabolism (three genes). Specific gene losses were also found in *D. hansenii* (eight genes missing), *K. lactis* (five genes

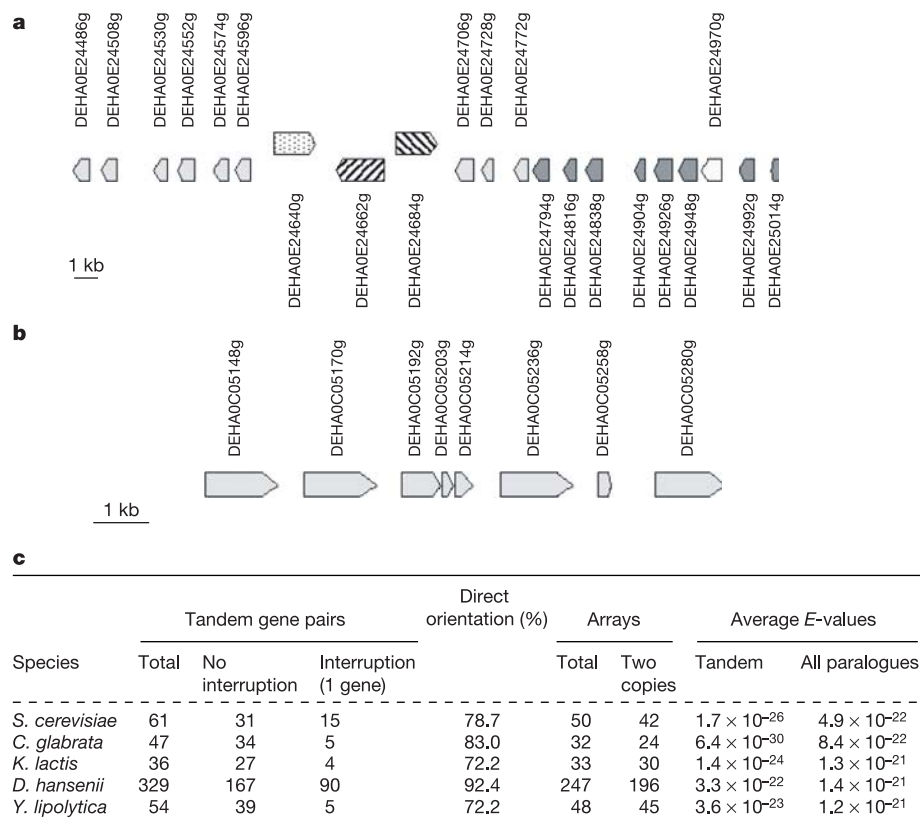


Figure 3 Tandem gene repeats. **a**, Example of two large tandem arrays (nine gene copies plus one fragment (light grey), and eight gene copies (dark grey)) forming a partially intermingled structure on one *D. hansenii* chromosome. Genes are of unknown function. The first array is interspersed by three other genes (dotted and striped). **b**, Example of a tandem array of four gene copies encoding NADH dehydrogenase, plus a pseudogene and

a gene fragment. **c**, Computation of all tandem gene arrays in the five yeast genomes (see Methods). The table indicates the total number of pairs of genes (column 2), either contiguous (column 3) or separated by one intervening gene (column 4). Column 5 indicates the proportion (%) of gene pairs in direct orientation.

missing) and *Y. lipolytica* (39 genes missing), but their functional coordination is less obvious.

Contrary to gene loss, the acquisition of new genes in a species raises the question of their origin. Horizontal gene transfer, which is quantitatively important in prokaryotes, is very rare in hemiascomycetes. A few examples of genes occurring in only one yeast species but having close homologues in Bacteria could, however, be detected (see Supplementary Table S11). This is the case for eight genes (including two pairs of paralogues) in *Y. lipolytica*, five genes (including one pair of paralogues) in *K. lactis* and one gene in *D. hansenii*. Many of them encode metabolic enzymes. No convincing case could be detected in *C. glabrata*. If these actually represent examples of horizontal gene transfer, this phenomenon accounts for less than 1% of the total gene number in yeasts. In addition, a small number of other genes seem to be species specific in the sense that they are limited to one of our five yeast species (as deduced from protein families; see Table 4) and have no significant homologue in general databases. The origins of these genes remain unclear and their function is generally unknown although, in many cases, they form paralogous families in the only species where they are found, indicating rapid expansion after their acquisition or *de novo* creation.

Finally, the broad evolutionary range covered by our sequenced yeast genomes allowed us to investigate the sequence conservation of proteins known to be involved in protein–protein interactions in *S. cerevisiae* compared to the average level of sequence conservation for all proteins. Previous comparisons with *S. pombe* and *C. elegans*

already indicated the significant conservation of such proteins²⁸. Systematic examination across the evolutionary span of hemiascomycetes (Table 5) shows that homologues of the set of 785 ‘interacting proteins’ of *S. cerevisiae* are more conserved than average (homologues to the entire *S. cerevisiae* protein set). This trend is even more pronounced for the homologues of the 1,535 proteins of the ‘complex set’ and becomes more apparent when the phylogenetic distances increase (compare ratios in *C. glabrata* and *Y. lipolytica*).

Discussion

The central strategy of this work was to examine eukaryotic genome evolution by confining the comparisons within a single phylum, while exploring its evolutionary range as widely as possible. Each species revealed unique signatures in its genome, reflecting the fact that distinct mechanisms have predominated in each phylogenetic branch. With its larger genome size, not linked to a significantly larger gene repertoire, the highly redundant genome of *Y. lipolytica* shows a strong tendency for map dispersion. This dispersion is visible at various levels: a near complete absence of duplicated blocks despite a high number of paralogous genes; a higher number of tRNA genes; a higher number of rDNA loci; a dispersion of the 5S RNA genes; and the specific duplication of other non-coding RNA genes. By contrast, the other yeast species show significant constraints on genome size, possibly associated with their ability to duplicate genes in an ordered manner as revealed by the presence of duplicated blocks and tandem gene repeats in their genomes. This

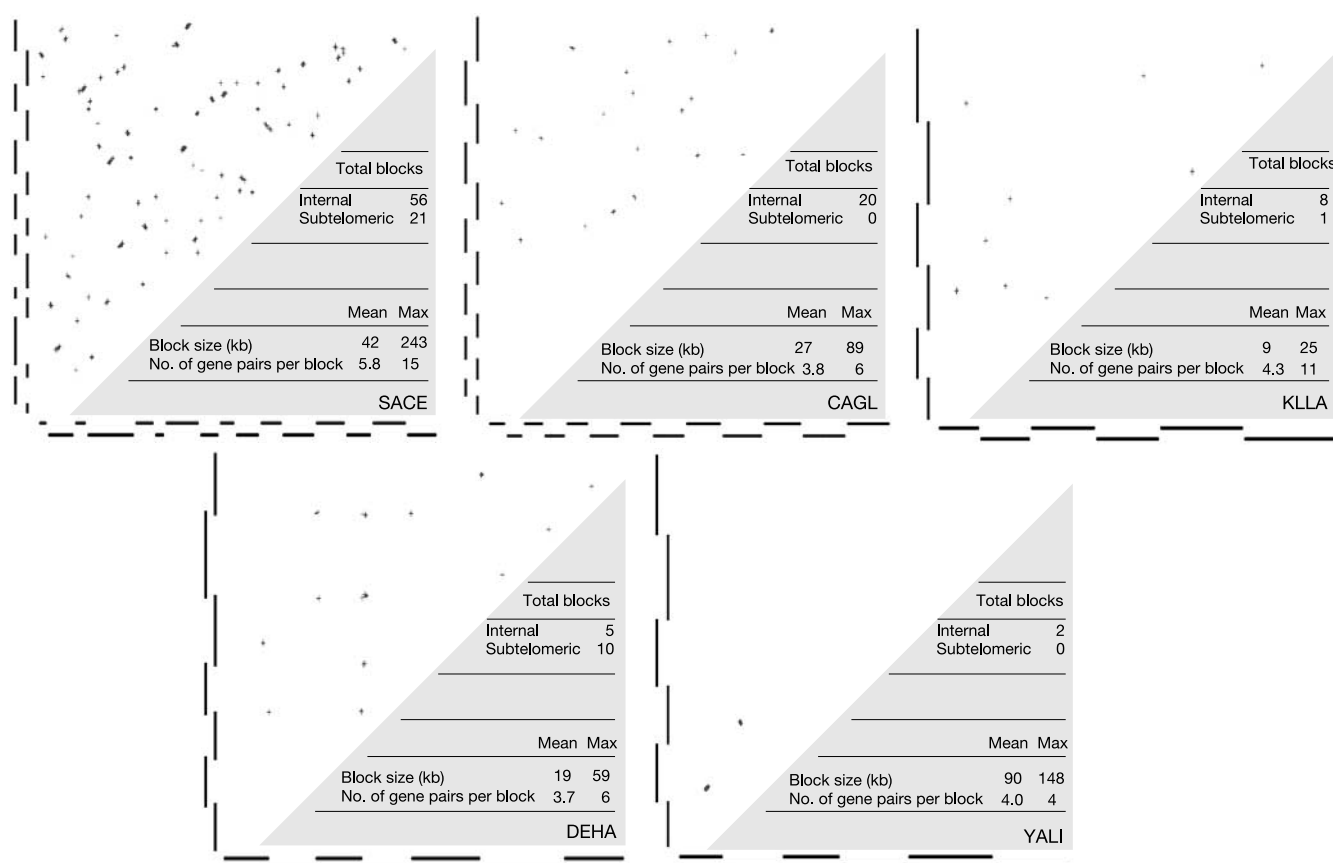


Figure 4 Detection of ancient duplicated blocks in each yeast genome. Chromosomal blocks containing at least three pairs of paralogous genes with a probability lower than 0.01 to occur by chance (see Methods) are plotted on two-dimensional diagrams each representing an entire yeast genome (chromosomes, represented by shifted bars, are

ordered by increasing numbers from the lower-left corner). Each cross on the plots corresponds to a pair of paralogous genes belonging to a defined block. The shaded triangle in each plot gives numerical descriptions of the duplicated blocks found for each species. Species abbreviations as in Fig. 1.

last trend is particularly acute in *D. hansenii* but exists in all species. The remaining three species have acquired novel features such as the triplication of the *MAT* cassettes and short centromeres. Compared with *K. lactis*, which has the shortest and least redundant genome of the five yeast species studied, *S. cerevisiae* and *C. glabrata* partly

share the traces of an extensive duplication event in their common ancestry (shared duplicated blocks). However, they present very different levels of duplication, which can only be reconciled with a single whole-genome duplication event, as proposed previously^{16,17,24}, if one assumes a significantly higher rate of loss of

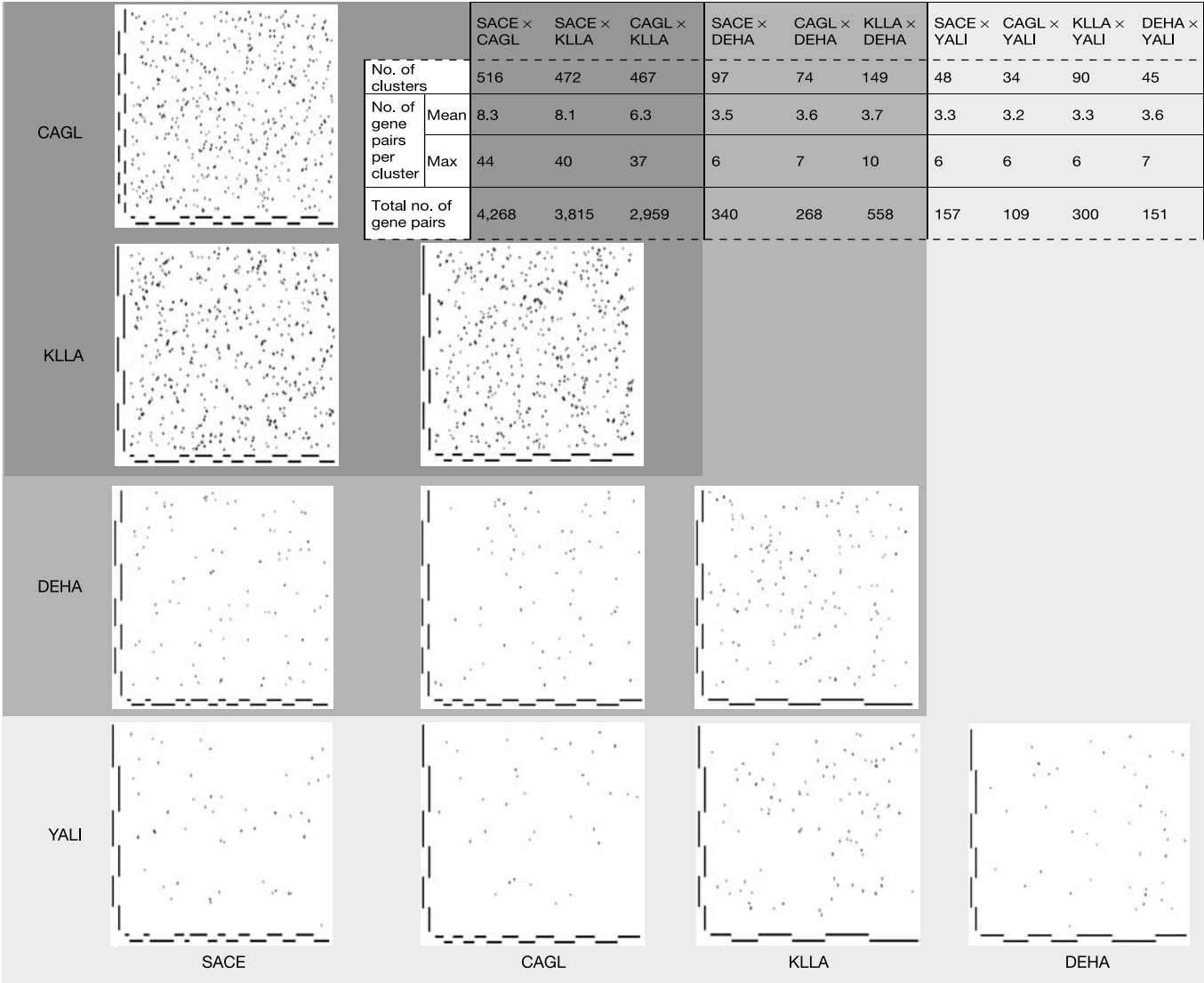


Figure 5 Conservation of synteny between yeast genomes. Syntenic clusters (defined in Methods) are indicated by crosses on all possible pairwise comparisons of the five yeast genomes (same representation of chromosomes as in Fig. 4). The dark grey background indicates comparisons between the three closest relatives (*S. cerevisiae*, *C. glabrata* and

K. lactis), medium grey indicates comparisons between these species and *D. hansenii*, and light grey indicates all comparisons including *Y. lipolytica*. The numerical characterization of syntenic clusters between any two yeast species is given in the inset (top). Species abbreviations as in Fig. 1.

Table 5 Sequence conservation in interacting proteins and multiprotein complexes

Protein sets	Total number of proteins in set	BLAST					ClustalW
		BDBH	S. cerevisiae proteins with a BDBH (%)				Alignment score
			C. glabrata	K. lactis	D. hansenii	Y. lipolytica	
S. cerevisiae set	5,807	2.62	74.1	73.0	63.8	51.5	13,587.4
Interacting set	785	3.26	89.8	87.9	80.3	68.0	14,351.9
Complex set	1,535	3.47	93.3	92.6	86.0	74.8	16,362.4

Experimentally defined protein–protein interactions in *S. cerevisiae* were taken from ref. 34. For each protein set of *S. cerevisiae*, sequence conservation in other genomes was estimated by bi-directional best hits (BDBH) in BLASTP alignments with *E*-value $\leq 10^{-8}$ (BLAST heading), and the average ClustalW alignment scores for *S. cerevisiae* proteins giving significant BDBHs with all four other genomes (ClustalW heading). The table indicates the average number (over each protein set) of the four yeast genomes giving a significant BDBH with *S. cerevisiae* proteins (column 3), and the percentage of proteins of each set giving a significant BDBH with each other yeast species (columns 4–7). The last column indicates the average ClustalW alignment scores for each set.

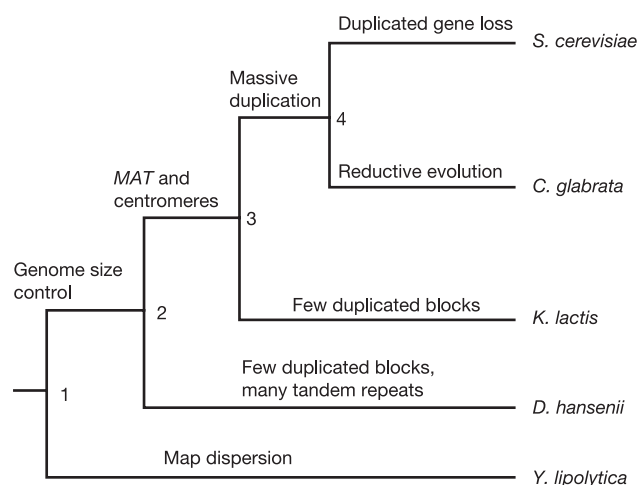


Figure 6 Major evolutionary events in the genomes of hemiascomycetes. Shown is a cartoon of the evolutionary history of the four sequenced yeasts, plus *S. cerevisiae*. The tree topology is based on 25S rDNA sequences (see Supplementary Fig. S17). The most conspicuous evolutionary signatures are summarized on each branch. The tendency for map dispersion of *Y. lipolytica* is visible at a variety of levels (see text). The other yeasts share an ability to duplicate genes in an ordered manner. An accidental whole-genome duplication event occurred in the common ancestor of *S. cerevisiae* and *C. glabrata*. It has been followed by extensive gene loss of paralogous copies.

the duplicated gene copies in the *C. glabrata* lineage than in *S. cerevisiae*. The reductive evolution of the *C. glabrata* genome was not anticipated and may be related to the adaptation of this yeast to life as a human pathogen.

Despite the conspicuous organizational differences between the five yeast genomes, all results are consistent with the topology of their phylogenetic tree (see Fig. 6). Accepting the risks inherent in the delicate exercise of reconstructing an evolutionary past in which accidental events may be misleading, we propose the following steps in the evolution of hemiascomycetes. At the separation between *Y. lipolytica* and the four other yeasts (node 1, Fig. 6), one branch lost the DNA transposons, kept the retrotransposons and responded to stronger genome size constraints (the nature of which remains to be elucidated) by simultaneously reducing its number of introns. In this context, the coordinated duplication of genes, either as large chromosome segments or as tandem repeats, became the major route to create the paralogous copies on which subsequent diversification is based. At the next separation (node 2), one branch, keeping a low level of segmental duplication, underwent extensive formation of tandem gene repeats, whereas the other, perhaps associated with the appearance of new centromeres providing for better chromosomal segregation, underwent segmental duplication to create the three *MAT* cassettes that changed the sexual capabilities and, consequently, the whole evolutionary future of species on this branch. At node 3, while some of these yeasts retained these features, leading to *K. lactis* (which can be regarded as representative of the ancestor of this branch), others underwent a whole-genome duplication event subsequently compensated by extensive gene loss among the newly formed paralogous pairs. The loss of paralogues sometimes leaves behind visible relics, as with *S. cerevisiae*²⁷. In *C. glabrata*, this loss of paralogues has been so extensive as to result in a reductive evolution with loss of function and a general degree of genome redundancy nearly equivalent to that of *K. lactis* (and significantly lower than that of *Y. lipolytica* and, even more so, *D. hansenii*).

The diversity of evolutionary routes taken by each branch was unexpected at the start of our work. However, it becomes less surprising if one considers that, despite their morphological and

biological similarities, the five hemiascomycete yeast species encompass an evolutionary span as large as the entire phylum of chordates. Relative to this broad evolutionary spectrum, the number of yeast species now entirely or partially sequenced remains limited. With over 700 species described²⁹, other interesting novelties may lie hidden in the genomes of other hemiascomycetes. Together with the fact that yeasts are such powerful experimental systems, this should stimulate future studies with impacts that reach far beyond yeasts. □

Methods

Sequencing and sequence assembly

Plasmid libraries (insert sizes 3–5 kb) were constructed using randomly sheared total DNA purified from the haploid type strain of each species and used for shotgun sequencing (Table 1). Bacterial artificial chromosome (BAC) libraries were used for sequence finishing, as required, and for assembly verification. Sequence electrophoresis and base calling were performed on either a 3700 Genetic Analyser (ABI PRISM BigDye Terminator) or a Licor 4200L DNA sequencer (dye primers). The traces were assembled using Phrap and the resulting contigs were checked and ordered using in-house software (Cover and Coverparse; M. Levy, unpublished). Gaps were closed by the primer-walking method. Regions of insufficient quality were resequenced using either new primers or a different chemistry. Correct assembly and collinearity of contigs and scaffolds were systematically (*C. glabrata*) or partially (*K. lactis*, *Y. lipolytica*) verified using BAC fingerprints.

Annotation

Genome annotation was conducted using the CAAT-Box system³⁰. Automatic predictions were manually curated and each open reading frame (ORF) was classified as coding (CDS), non-coding (false ORF) or pseudogene according to Génolevures annotation standards (<http://cbi.labri.fr/Genolevures>). The codon usage matrix used to screen for ORFs was trained on a sample of coding and non-coding sequences, initially produced by automatic identification, and subsequently using curated CDS. Annotation was performed iteratively across successive sequence assemblies (each file is conserved if no sequence edition occurs, and it is replaced otherwise). In a final stage, 291 small ORFs missed by the automatic procedures were identified with the help of the protein family classification (each protein of incomplete families was used as a tBLASTn query against all remaining intergenic sequences). Introns were identified from consensus splice sites and branch points, and from comparisons to *S. cerevisiae*³¹. Genes encoding tRNA molecules were identified according to ref. 18 (see Supplementary Table S3). All other non-coding RNA molecules (Supplementary Table S6), rRNAs, transposons, centromeres (Supplementary Table S1) and telomeric repeats (Supplementary Table S2) were identified by sequence comparisons and manual curation.

All annotated genetic elements were designated using a new nomenclature system (<http://cbi.labri.fr/Genolevures>). Briefly, elements are numbered serially along each sequence contig or scaffold from the left to right of each chromosome using 11 incremental steps (to limit errors and offer the possibility for subsequent insertion of newly recognized elements). The element nomenclature indicates the species (four letters), the project or strain number (one numeral), the chromosome (one letter) followed by the serial number (for example, CAGL0G08492g). The suffix identifies the type of element ('g' stands for any element whose RNA product may be translated by the genetic code; 'r' for elements whose RNA product is not translated; 's' for a *cis*-acting element; and 'v' for intergenes (intervening)).

Classification of proteins into families

A data set of 34,824 amino acid sequences corresponding to all annotated and predicted proteins (not necessarily subsequently confirmed) from our four yeast species plus the annotated proteins of *S. cerevisiae* was used to construct protein families. Two sets of sequence alignments were produced separately using Smith–Waterman and BLAST algorithms. After filtration for statistical significance, the pairs considered as homologous were clustered using the MCL method³² with a variety of inflation parameters (see Supplementary S12). Results of clusterings applied separately on the Smith–Waterman and BLAST pairs were compared using a graph-based technique producing a best coincidence of 3,016 strictly identical protein families (exactly the same set of proteins using the two alignment methods). Another set of 394 families, corresponding to exactly the same set of proteins split into two families (or merged into a single one) when comparing the two methods, was easily reconciled using motif search programs. The two sets form the 'robust' families. Reconciliation was not attempted for the more complex cases in which the same set of proteins was classified into partially overlapping families by each of the two methods. Instead, the union of all such proteins was retained as 'consensus' families (a total of 1,311 families).

Identification of duplicated blocks and syntenic clusters

Using homologous gene pairs as defined from protein families, maps were compared with the ADHoRe program³³ (r^2 cutoff = 0.8, maximum gap size = 35 genes, minimum number of pairs = 3). Results were filtered such that the maximum probability for a segment to be generated by chance was lower than 0.01. Paralogous pairs were used to define ancient duplicated blocks within each genome (Fig. 4), whereas orthologous pairs were used to define conserved syntenic clusters between two genomes (Fig. 5).

Identification of tandem arrays

Each genome was systematically examined for the presence of repeated gene arrays. Arrays are defined as a succession of genes encoding paralogous proteins (as defined by protein families and with a BLASTP *E*-value $<10^{-20}$) allowing a maximum of ten intervening genes between two closest members of an array. Arrays were defined regardless of gene orientation.

Received 3 February; accepted 19 April 2004; doi:10.1038/nature02579.

- Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 563–567 (1996).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Amanatides, P. G. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Holt, R. A. *et al.* The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129–149 (2002).
- Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
- Stein, L. D. *et al.* The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biol.* **1**, 166–192 (2003).
- Stephen, A. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
- Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
- Galagan, J. E. *et al.* The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 857–868 (2003).
- Wood, V. *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
- Souciet, J. L. *et al.* Genomic exploration of the hemiascomycetous yeasts. *FEBS Lett.* **487**, 3–147 (2000).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
- Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**, 304–307 (2004).
- Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
- Marck, C. & Grosjean, H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189–1232 (2002).
- Soonhammer, E. L. & Koonin, E. V. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* **18**, 619–620 (2002).
- Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167 (2002).

- Brown, C. J., Todd, K. M. & Rosenzweig, R. F. Multiple duplications of yeast hexose transport gene in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* **15**, 931–942 (1998).
- Lalo, D., Stettler, S., Mariotte, S., Slonimski, P. P. & Thuriaux, P. Two yeast chromosomes are related by a fossil duplication of their centromeric regions. *C. R. Acad. Sci. III* **316**, 367–373 (1993).
- Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
- Wong, S., Butler, G. & Wolfe, K. H. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc. Natl Acad. Sci. USA* **99**, 9272–9277 (2002).
- Koszul, R., Caburet, S., Dujon, B. & Fischer, G. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**, 234–243 (2004).
- Fischer, G., Neugebäude, C., Durrens, P., Gaillardin, C. & Dujon, B. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* **11**, 2009–2019 (2001).
- Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. Evolutionary rate in the protein interaction network. *Science* **296**, 750–752 (2002).
- Kurtzman, C. P. & Fell, J. W. *The Yeasts, a Taxonomic Study* 4th edn (Elsevier, Amsterdam, 1998).
- Frangul, L. *et al.* CAAT-Box. Contigs-Assembly and Annotation Tool-Box for genome sequencing projects. *Bioinformatics* **20**, 790–797 (2004).
- Bon, E. *et al.* Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* **31**, 1121–1135 (2003).
- Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
- Vandepoel, K., Saey, Y., Simillion, C., Raes, J. & Van De Peer, Y. The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity between *Arabidopsis* and rice. *Genome Res.* **12**, 1792–1801 (2002).
- Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**, 248–250 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Tichit, S. Duthoy, S. Ferry and N. Zidane for technical assistance; A. Louis for help and expertise in the use of the INFOBIOGEN computational facilities; O. Jaillon, H. Roest-Crolius and their colleagues for sharing unpublished results on *Tetraodon nigroviridis*; and A. Goffeau, H. Feldmann, A. Nicolas, N. Huu-Vang, J.-P. Latgé, I. Moszer and M. Vergassola for discussions and advice. This work was supported by the Consortium National de Recherche en Génomique (to Génoscope and to Institut Pasteur Génomique), the CNRS (GDR 2354, Génolevures sequencing consortium) and the 'Conseil Régional d'Aquitaine'. B.D. is a member of Institut Universitaire de France.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.-L.S. (souciet@gem.u-strasbg.fr) or B.D. (bduron@pasteur.fr). Sequences have been deposited in EMBL under accession numbers CR380947–CR380959 for *C. glabrata*, CR382121–CR382126 for *K. lactis*, CR382127–CR382132 for *Y. lipolytica* and CR382133–CR382139 for *D. hansenii*.

Crystal structure of a self-splicing group I intron with both exons

Peter L. Adams, Mary R. Stahley, Anne B. Kosek, Jimin Wang & Scott A. Strobel

Department of Molecular Biophysics and Biochemistry, 260 Whitney Avenue, Yale University, New Haven, Connecticut 06520-8114, USA

The discovery of the RNA self-splicing group I intron provided the first demonstration that not all enzymes are proteins. Here we report the X-ray crystal structure (3.1-Å resolution) of a complete group I bacterial intron in complex with both the 5'- and the 3'-exons. This complex corresponds to the splicing intermediate before the exon ligation step. It reveals how the intron uses structurally unprecedented RNA motifs to select the 5'- and 3'-splice sites. The 5'-exon's 3'-OH is positioned for inline nucleophilic attack on the conformationally constrained scissile phosphate at the intron-3'-exon junction. Six phosphates from three disparate RNA strands converge to coordinate two metal ions that are asymmetrically positioned on opposing sides of the reactive phosphate. This structure represents the first splicing complex to include a complete intron, both exons and an organized active site occupied with metal ions.

RNA splicing involves the removal of an intron from a primary RNA transcript with the concomitant union of two exons¹. Most introns are removed by the spliceosome, a large ribonucleoprotein complex, but a subset of introns catalyse their own excision from the primary transcript, without the addition of accessory factors. The group I self-splicing intron from the ciliate *Tetrahymena* was the first of several catalytic RNAs to be discovered^{2,3}. This remarkable discovery revealed that not all enzymes are proteins.

Here we report the X-ray crystal structure (3.1-Å resolution, Table 1) of the group I self-splicing intron from the pre-transfer RNA^{lle} anticodon loop of the purple bacterium *Azoarcus* sp. *BH72* (ref. 4). This structure of the intron in complex with both of its exons (222-nucleotides) was crystallized in the state preceding the second step of splicing (pre-2S). Two previous X-ray crystal structures revealed the architecture of intron domains, but the exons and other helical elements were omitted in these structures^{5,6}. In the pre-2S complex, the 5'-exon has been cleaved, but is base-paired to the intron's internal guide sequence (IGS) via helix P1 (Fig. 1a). The 3'-exon remains covalently connected to the intron and also base pairs with the IGS via helix P10 (Fig. 1a). The terminal nucleotide of the intron, ΩG206 (ΩG), occupies the G-binding site.

The pre-2S state was captured by introducing 2'-deoxy substitutions at four positions: U - 1, A + 1, A205 and ΩG. These substitutions reduce the splicing activity a million-fold and eliminate cryptic splicing^{7,8}. The complex was prepared by annealing an RNA transcript (UP62; residues αG1-A190) with two RNA-DNA chimaeric oligonucleotides (CAT, 3 nt; and dCIRC, 22 nt) (Fig. 1a). The all-ribose version of this complex produced ligated exons, whereas the four deoxy-substituted complex showed no detectable activity after several days (Supplementary Information). The intron was co-crystallized with the RNA binding protein U1A (Fig. 1, see Supplementary Information)^{9,10}.

Intron architecture and substrate alignment

The architecture of the pre-2S intron is dominated by three coaxially stacked helical elements: P10, P1, and P2 (orange/yellow); P5, P4 and P6 (green); and P9.0, P7, P3 and P8 (blue) (Fig. 1b). These helices are aligned by extensive tertiary interactions that are mediated primarily by the joiner segments between the helices and 18 metal ions. As predicted, the peripheral regions of the intron are stabilized by two GAAA tetraloop-tetraloop receptor contacts, one between the P2 tetraloop and its receptor in P8, and a second between the P9 tetraloop and P5 (Fig. 1)¹¹. The relative placement of the P4-P6 and P3-P9.0 helices approximately matches what was

observed in the *Tetrahymena* apoenzyme structure⁶. However, the active site, which could previously only be inferred through model building^{6,12,13}, is ordered and appears to be in a catalytically relevant conformation within this pre-2S structure.

Collectively, base pairing and tertiary contacts align the exons and the ΩG for reaction. The 5'-exon's 3'-OH is positioned for inline nucleophilic attack on the conformationally constrained scissile phosphate at the intron-3'-exon junction. There is a complete reversal of strand direction at the 3' splice site between ΩdG of the intron and dA + 1 of the 3'-exon (Fig. 1, 2a). The 5'-exon O3' is within 3.5 Å of the scissile phosphate, and the angle of approach is consistent with the geometry expected for inline nucleophilic attack, resulting in inversion of stereochemistry¹⁴. Substrate alignment by Watson-Crick and non-canonical base pairings plays an important catalytic role in the peptidyl transferase and hairpin ribozymes^{15,16}. This structure clearly demonstrates that the group I intron also utilizes substrate alignment as a fundamental strategy to promote exon ligation.

3'-splice site selection and ΩG binding

The 5' and 3' splice sites must be accurately selected from among hundreds of nucleotides in the primary transcript in order to prevent insertions or deletions in the spliced product. On the basis of the pre-2S structure, the 3'-splice site is selected primarily by interactions with the ΩG (Fig. 2). There are no tertiary contacts to the 3'-exon and only a single tertiary hydrogen bond to the P10 portion of the IGS. ΩdG binds into an intimate pocket created by four consecutive nucleotides, A127-G130, at the junction of J6/7 and P7 (Fig. 2). The first residue (A127) stacks under ΩdG, and the second (G128) stacks over it (Fig. 2a). The fourth residue (G130) is coplanar with ΩdG and makes a Hoogsteen base pair with it. The Hoogsteen base pair to G130 is equivalent to that predicted on the basis of compensatory mutational analysis (Fig. 2c)¹⁷.

The conformation of the four residues in the G-binding motif is novel. Unlike previous models^{6,12}, A129 is at the P7 apex and forms a coplanar base triple with G128 (Fig. 2b). J6/7 residues U126-G128 adopt a maximally extended conformation that spans 22 Å across the breadth of the active site, compared to a typical distance of ~9 Å for equivalent functional groups within a stacked duplex (Fig. 2a). We suggest that the tautness of J6/7 may be critical for intron function in two ways: (1) it displaces G128 over the P7 major groove and pulls A127 deep into it, which allows ΩG to bind between them; and (2) the extended backbones of A127 and G128 serve as ligands for an active site metal ion (see below). This may explain why the

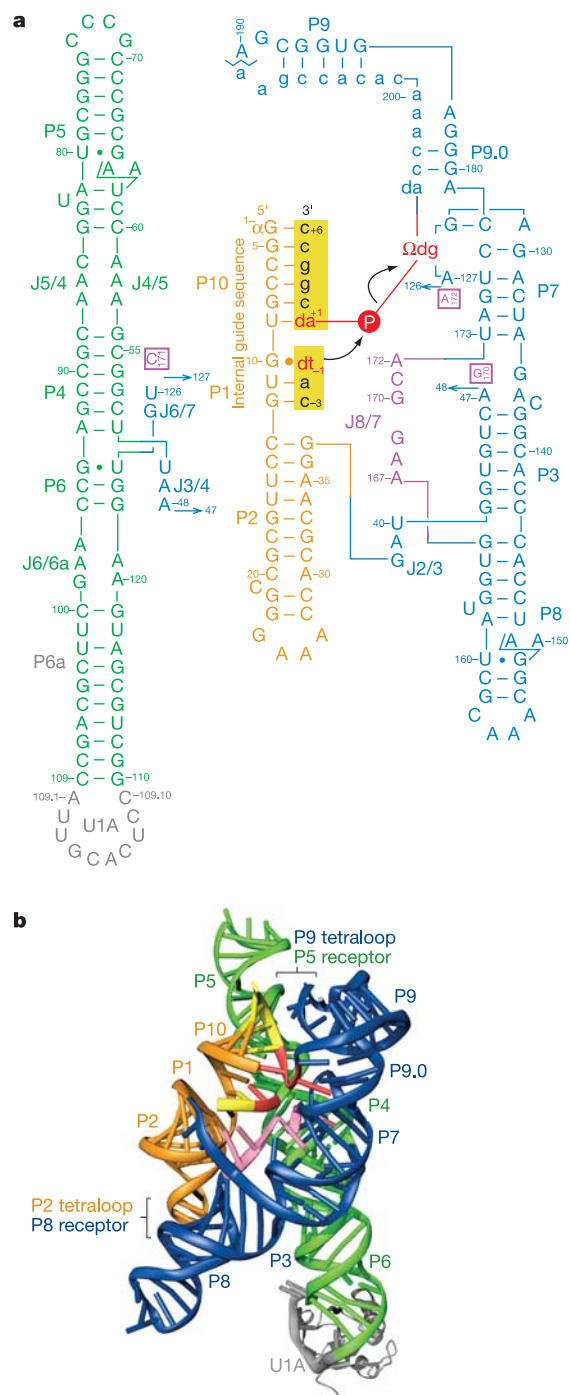


Figure 1 Overall secondary and tertiary structure of the *Azoarcus* pre-2S intron splicing complex. **a**, Intron secondary structure. The intron sequences, exon sequences and structural elements (P and J elements) are depicted in the colours used for all subsequent figures. The ligation reaction catalysed by this complex is shown by black arrows. The RNA transcript (UP62) comprising the majority of the intron is shown with capital letters, while residues derived from the two chimaeric oligonucleotides dCIRC (intron/3'-exon segment) and CAT (5'-exon segment) are shown in lower-case letters. The break between dCIRC and UP62 is indicated with a jagged line. Intron numbering follows the canonical nomenclature established for group I introns, in which the α G added in the first step of splicing is numbered 1. **b**, Overall structure of the intron. The backbone is depicted with a ribbon and individual bases depicted as cylinders. The two tetraloop-tetraloop receptor interactions on the periphery of the intron are indicated. This and all subsequent structure figures were prepared using RIBBONS⁵⁰.

length, but not the sequence of J6/7, is absolutely conserved¹⁸. It is reasonable to expect that every group I intron will adopt a similar binding motif for Ω G; however, on the basis of PRIMOS analysis¹⁹, this conformation is without precedent among any of the previously determined RNA structures.

5'-splice site selection

The crystal structure reveals the extensive network of tertiary interactions that the intron uses to select the 5'-splice site. The 5'-exon, which in this intron makes three base pairs with the IGS, is covalently disconnected from the intron after the first step of splicing. As a result, these tertiary contacts are critical for retaining the 5'-exon in the complex before ligation^{20–22}. Helices P1 and P2 are recognized exclusively in the minor groove through interactions with three joiner regions of the intron, J4/5–J5/4, J8/7 and J2/3 (Figs 1a and 3a). The functional groups within P1–P2 that were implicated as being functionally important using interference analysis^{21,22} correspond precisely with those that participate in tertiary interactions. This unambiguous convergence of structure and function is a strong indication that the crystallized RNA is in a conformation that is relevant for splicing.

The 5'-splice site is specified by an invariant wobble mismatch between U-1 at the end of the exon and a G within the IGS (Fig. 3bc)¹⁸. The G-U pair is recognized by the conserved A-rich J4/5–J5/4 symmetric loop, using the wobble receptor motif. The conserved As in the loop are mispaired such that one A from each strand (J4/5 residue A58 and J5/4 residue A87) is presented on the minor groove surface to contact the wobble pair (Fig. 3c)^{5,23}. This interaction defines a simple scheme for RNA helix packing, namely that a G-U pair in one helix is complementary to two consecutive A-A mismatches in the second helix. Modelling of the U-1 O2' into the active site shows it to be within hydrogen-bonding distance of the A87 2'-OH. An equivalent hydrogen-bonding network in the *Tetrahymena* intron contributes approximately 1,000-fold to the chemical step of catalysis (Fig. 3c)^{24–26}. Although the wobble receptor appears to be used for 5'-splice site selection among virtually all group I introns¹⁸, there are no examples of this motif in either the 30S or 50S ribosomal RNA^{27,28}. Possibly, there is selection against this motif because it may render the linkage 3' of the U more labile.

Active-site metal ions

The group I intron is an obligate metalloenzyme expected from biochemical studies to have at least three magnesium ions within its active site^{29,30}. These metals, termed M_A, M_B and M_C, were investigated by metal-ion specificity-switch experiments in which sulphur or amino substitutions were rescued with soft divalent metal ions^{30–34}. In the exon ligation reaction, these metals have been proposed to activate the nucleophile (M_A)³¹, stabilize the leaving group (M_B)³², neutralize the negative charge on the trigonal bipyramidal phosphorous transition state (M_A and M_C)^{30,33}, and coordinate to the Ω G O2' (M_C)³⁴.

The overall RNA architecture results in extremely high phosphate density in the active site. Six phosphates from J5/4, J6/7 and J8/7 converge to coordinate two metals, identified here as M₁ and M₂, on opposing sides of the scissile phosphate. The two metal ions are positioned asymmetrically relative to the scissile phosphate, with M₁ closer than M₂ to the critical pro-R_p non-bridging oxygen¹⁴. The two metals are 5.4 Å apart from each other. Such a distance is larger than the 3.9-Å distance predicted for the intron on the basis of protein-based phosphotransferases³⁵. There is no structural evidence for the third active-site metal ion predicted to be near the Ω G O3' (M_B)³², and there are no obvious ligands available to coordinate with this metal. However, M_B had the lowest affinity in studies on the *Tetrahymena* ribozyme²⁹, and we cannot rule out the possibility that M_B is disordered in this pre-2S complex.

The identity of the two active-site metals was investigated by

soaking the crystals with heavy metal ions that mimic physiological cations and have anomalous scattering (Fig. 4a). We assigned the M_1 as a Mg^{2+} on the basis of a soak with Yb^{3+} and Mn^{2+} , metal ions that mimic Mg^{2+} . We observed Yb^{3+} and Mn^{2+} binding to a single position within the intron, coincident with the M_1 binding site (Fig. 4a). Within the native structure, the 2.0–2.4 Å distances of ligands to the metal and the refined temperature factors are also

consistent with this assignment. The coordination environment of M_1 is rather unusual. It appears to be directly coordinated to the non-bridging oxygens of four phosphates and outer sphere coordinated to a fifth (Fig. 4b). Only one Mg^{2+} ion (M_{33}) among 116 examples in the refined 50S ribosome structure produces inner-sphere coordination to four or more phosphates (D. J. Klein & T.A. Steitz, personal communi-

Table 1 Statistics for X-ray crystallographic structure determination of the pre-2S intron complex

Space group	P4 ₁ 22	
Cell dimension	$a = b = 108.54 \text{ \AA}$, $c = 249.16 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$	
Protein Data Base accession code	1T42	
Native data		
Resolution (Å)	50–3.1	
R_{merge} (%)	9.8 (90)†	
Intensity ratio (I/σ)	17.8 (1.8)†	
Number of reflections	27,894	
Derivatives*	Yb ³⁺ (SIR/AS)	Tl ⁺ (MAD)
Resolution (Å)	4.8	4.1
R_{merge} (completeness) (%)	11 (99)	8.5 (98)
Figure of merit	0.33 (at 5 Å)	0.37
Phasing power (acentric/centric)	0.91/0.72	0.52/0.51
R_{cullis} (acentric/centric)	0.89/0.81	0.95/0.85
Number of heavy-atom sites	1	13
Structure refinement		
Resolution (Å)	50–3.1 (3.29–3.10)†	
Number of reflections, $F > 0$; percentage	24,640; 88.3%	
Number of test reflections; percentage	1,086; 3.9%	
R_{free} (%)	30.4 (40.3)	
R_{working} (%)	26.1 (33.7)	
Estimated coordinate error (Luzzati and σA methods)	0.44–0.46 Å	
r.m.s.d. (bond)‡	0.010 Å	
r.m.s.d. (angle)‡	1.5°	

*The reported statistics included only two of the eleven derivative data sets used to generate interpretable electron-density maps.
†Statistics for the highest-resolution shell are in parentheses.
‡Root-mean-square deviation (r.m.s.d.) from ideal bond lengths and angles.

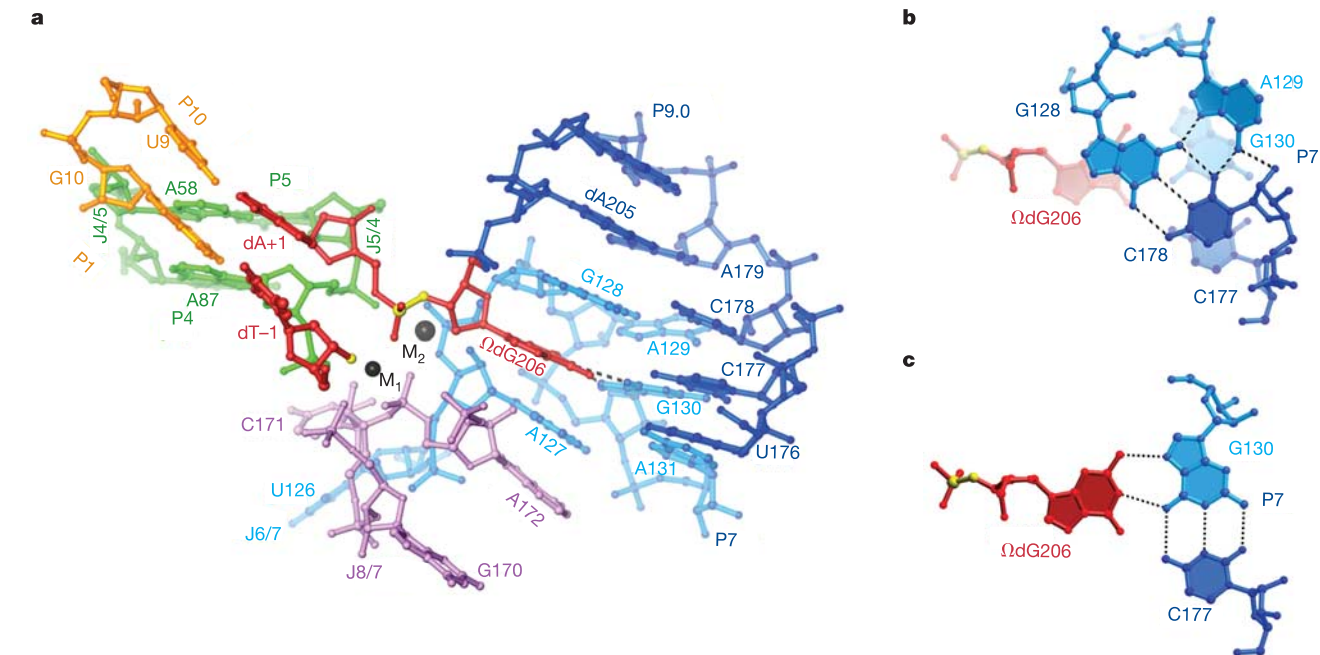


Figure 2 The G binding motif and the structure of the pre-2S active site. The nucleophile, scissile phosphorus, leaving group and cleaved bond are shown in yellow. **a**, Side view of the active site oriented to emphasize the stacking interactions that mediate ΩdG206 binding. The two active-site metal ions are shown as black spheres. **b**, Top-down view of

the base triple between the unpaired A129 and the closing G128–C178 base pair in the P7 helix. **c**, Top-down view of the base triple between ΩdG206 and the G130–C177 base pair.

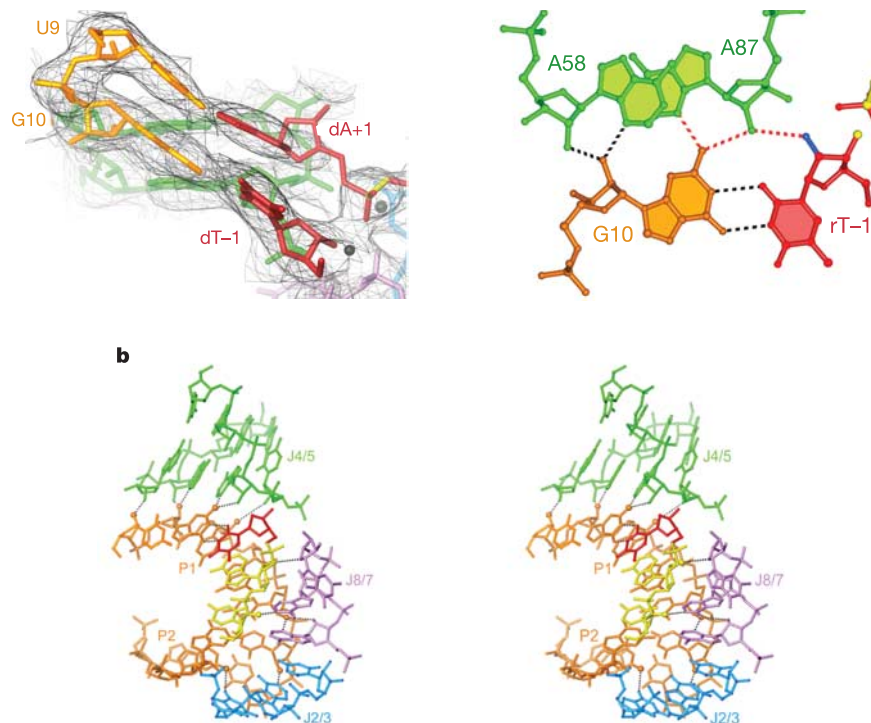


Figure 3 Recognition of the P1 substrate helix by the active site. **a**, Experimental electron density contoured at 1σ of the G-U wobble pair docked into J5/4–J4/5. The density for the active-site metal ions is also visible. **b**, Stereo view of minor-groove-mediated P1–P2 helix docking. Enlarged atoms correspond to the functional groups identified by

interference analysis as being important for ribozyme function^{21,22}. **c**, Wobble–wobble receptor motif that specifies the 5′-splice site. The U-1 2′-OH (blue) is modelled into the structure with the predicted network of transition state stabilizing hydrogen bonds in red. The scissile phosphate is also shown.

cation). Most importantly, metal M_1 is coordinated to both exons via the U-1 O3′, the reaction nucleophile, and the pro- R_p oxygen of the scissile phosphate (Fig. 4b). All of the ligands are completely consistent with the identification of M_1 as the biochemically defined Mg^{2+} ion M_A ^{31,33,36}.

Using similar criteria, M_2 was found to be a K^+ within the unreactive Ω dG206 complex. Neither Yb^{3+} nor Mn^{2+} bound to the M_2 site under the conditions tested (Fig. 4a and data not shown); however, M_2 was the primary binding site for thallium (Tl^+), which is a heavy-metal mimic of K^+ (ref. 37 and references therein). Temperature factor refinement and the somewhat longer bond lengths of M_2 to its ligands are also consistent with the assignment as a K^+ . The K^+ ligands include three non-bridging phosphate oxygens and two water molecules (W_1 and W_2). One of the waters (W_1) is hydrogen-bonded to the pro- S_p oxygen of the scissile phosphate and is close to the leaving group Ω dG206 O3′ (Fig. 4b).

The crystallographic observation of an active site K^+ ion is unexpected because the *Azoarcus* and other group I introns do not explicitly require K^+ for reactivity, and K^+ can be inhibitory at high concentrations^{38–40}. To explore the possibility that K^+ binding to M_2 may be a consequence of deleting the Ω G206 2′-OH group, which may serve as one of the M_2 ligands, we undertook crystallization of a complex with a ribose at all positions within the intron and 3′-exon. This complex can catalyse the exon-splicing reaction, albeit at a rate reduced 10³-fold by a single 2′-deoxy substitution at U-1 (Supplementary Information). The reduced diffraction quality of this reactive complex (3.7-Å resolution) precluded crystallographic refinement, but heavy-metal soaks could still be used to explore the identity of the active site metal ions. Yb^{3+} bound exclusively to M_1 , consistent with its identification as a Mg^{2+} . Strikingly, no density was observed for Tl^+ binding at M_2 , although Tl^+ bound to equivalent sites elsewhere within the intron (Sup-

plementary Information). Electron density persisted at M_2 , but the identity of the metal (Mg^{2+} or K^+) could not be unambiguously assigned at this resolution. The absence of significant peaks in an $F_{obs(ribo)} - F_{obs(deoxy)}$ map implies that there are no gross conformational changes within the active site upon introduction of the Ω G 2′-OH.

The Ω G 2′-OH is likely to be a ligand for M_2 on the basis of modelling of the Ω G206 2′-OH into the deoxy complex. Such coordination is predicted for metal M_C , a catalytic metal ion established biochemically to be a Mg^{2+} (refs 34, 41). This suggests that deletion of Ω G 2′-OH converted M_2 from a Mg^{2+} site into a K^+ site. Mg^{2+} binding to M_2 makes it necessary that the active site be reorganized slightly to reduce the bond lengths and number of ligands to the metal. K^+ is substantially less capable than Mg^{2+} at modulating the pK_a of its ligands (Mg^{2+} aqua ion, 11.4; K^+ aqua ion, 16), which may explain why the single deoxy substitution at Ω G has such a dramatic inhibitory effect on activity (reduced 10⁶-fold).

Implications for catalysis

Using the structure and the extensive biochemical literature on group I introns, we asked how the intron promotes the exon-splicing reaction. The Mg^{2+} at M_1 is well positioned to serve as a Lewis acid in the activation of the U-1 O3′ oxyanion and electrostatic stabilization of the developing negative charge on the scissile phosphate^{30,31,35}. The structural observations for M_1 and transition-state expectations for M_A seem to be in complete agreement. The hydrogen-bonding network between the hydroxyls of U-1 and A87 and the amine of G10 also activates the O3′ for nucleophilic attack²⁶.

The Mg^{2+} ion at M_1 is analogous to metal M_A in all RNA and DNA polymerases⁴². In these cases, the metal ion is coordinated to the carboxylate side chains of two or three amino acids in addition

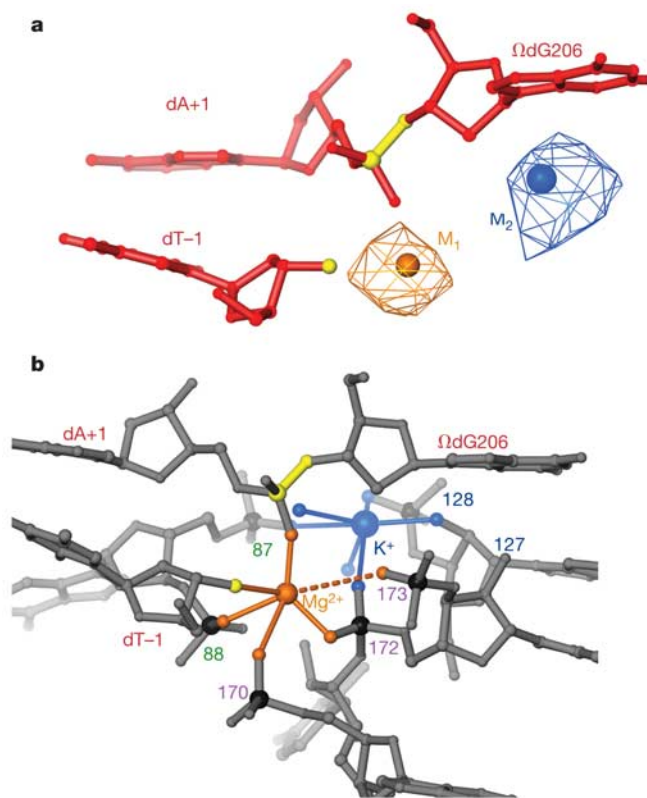
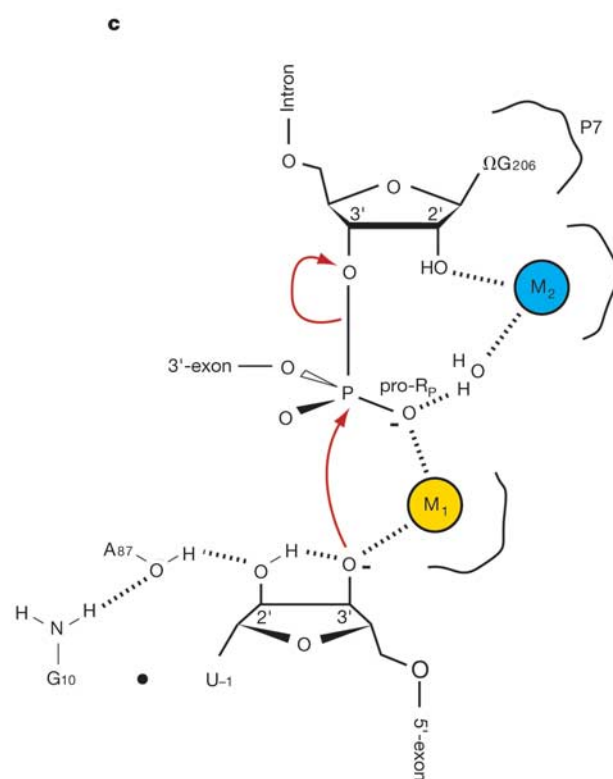


Figure 4 Active-site metal ions and their ligands. **a**, Identity of the metals based upon heavy-metal soaks. The two anomalous X-ray scattering electron-density maps for Yb^{3+} (orange) and Ti^{3+} (blue) are overlaid on the active site and contoured at 30σ and 33σ , respectively. The locations of the metal ions in the native structure are shown as solid spheres. **b**, Coordination of the active-site metal ions. The Mg^{2+} ion (M_1) and its ligands are shown in orange. The K^+ ion (M_2) and its ligands are depicted in blue. The

to the non-bridging oxygen of the scissile phosphate. Thus, the RNA intron is able to use its phosphate backbone to coordinate an active-site metal ion similarly to the way a protein uses the side chains of aspartates and/or glutamates.

The role of M_2 is also in reasonable agreement with the biochemical observations for metal M_C with one caveat. M_C was predicted to directly coordinate both the ΩG 2'-OH and the pro- R_P oxygen of the scissile phosphate^{30,34,41}. The second interaction seems to be too remote for M_2 (5.0 Å) in the ground-state structure. It is possible that the Mg^{2+} moves to achieve direct coordination in the transition state of the active complex, but M_2 may have its primary effect through direct interaction with the ΩG 2'-OH and a water molecule that bridges to the scissile phosphate. This metal could stabilize the transition state by indirectly coordinating the scissile phosphate's non-bridging pro- R_P oxygen (Fig. 4c). M_2 may also act as a Lewis acid to activate the ΩG 2'-OH, or alternatively to activate the bound water molecule. The latter scenario is intriguing because W_1 is reasonably close to the leaving group $\Omega\text{G}206$ O3' (2.7 Å). Activation by M_2 may allow the water to function as a general acid for protonation of the O3'. These possibilities await biochemical investigation, but the pre-2S intron structure provides a stereochemical context for these mechanistic considerations.

The reaction for the second step of exon ligation is chemically equivalent for the self-splicing group I and group II introns and the spliceosome¹. In each system the reaction involves displacement of the intron's O3' from the scissile phosphate by the 5'-exon's O3'. All three of these reactions require precise alignment of the 5' and 3'-



nucleophile, the scissile phosphorous, the leaving group and the labile bond are coloured yellow. All other residues are coloured grey, except for the phosphorus atoms, which are black. The residue labels are coloured according to the scheme in Fig. 1a. **c**, Proposed reaction mechanism for the second step of splicing by the bacterial intron. M_1 and M_2 are probably both Mg^{2+} ions.

exons within the complex, and all three are likely to be catalysed by RNA using metal-ion cofactors^{43,44}. As a result, the chemical themes of splice site selection, exon alignment and catalytic metal-ion positioning, which are manifest in this group I intron splicing complex, are likely to find parallels in group II and pre-messenger RNA splicing. □

Methods

Complex formation and crystallization

The UP62 RNA was prepared by T7 RNA polymerase transcription *in vitro*. UP62 was fused at its 3'-end to the antigenomic HDV ribozyme, which was processed post-transcriptionally to generate a homogeneous 3'-terminus containing a 2'-3'-cyclic phosphate. Chimeric oligonucleotides were purchased from Dharmacon Research, deprotected, dried and used in crystallization trials without further purification. The U1A recognition sequence was placed at the end of P6a and the intron was active with this modification (Supplementary Fig. 2). The U1A protein containing a double (Y31H and Q36R) mutation, along with the modified selenomethionine U1A protein, were expressed and purified essentially as described¹⁰. RNA (120 μM) was heated to 50 °C in folding buffer (10 mM sodium cacodylate pH 6.8, 10 mM KCl and 10 mM MgCl_2) and allowed to cool before the addition of the two chimeric oligonucleotides (dCIRC and CAT, 150 μM each) and the U1A protein (100 μM). Two volumes of the complex were combined with one volume of 26% 2-methyl-2,4-pentanediol (MPD), 50 mM sodium cacodylate pH 6.8, 40 mM MgCl_2 and 0.2 mM $\text{Co}(\text{NH}_3)_6^{3+}$ at 25 °C. Crystals were grown by the hanging-drop vapour diffusion method. They grew to a maximal size of $500 \times 200 \times 200$ μm in approximately 14 days. Crystals were stabilized in a solution containing 30% MPD, 50 mM sodium cacodylate pH 6.8, 40 mM MgCl_2 and 0.2 mM $\text{Co}(\text{NH}_3)_6^{3+}$ before flash-freezing. The complex with ribose at ΩG and neighbouring positions (rCIRC) was annealed and crystallized in the same manner, although the crystals were more fragile and had to be harvested and frozen within four days.

Structure determination

The crystal structure of the pre-2S complex was solved at 3.1-Å resolution using single isomorphous replacement with anomalous scattering (SIRAS), single-wavelength

anomalous dispersion (SAD) and multiwavelength anomalous dispersion (MAD) experiments using a combination of derivatives (Table 1). The structure was refined using CNS⁴⁵. Native and derivative data were collected at 100 K on beamline X25 at the National Synchrotron Light Source and processed using the HKL package⁴⁶. The pre-2S complex crystallized in the space group P4₁22 ($a = b = 108.54 \text{ \AA}$, $c = 249.16 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$) with one molecule in the asymmetric unit (solvent content, 72%). Phasing information was obtained using anomalous X-ray scattering data collected from several derivatives, including: one Yb³⁺ data set (soaked at 0.5 μM , diffraction to 4.5 $\text{\AA}$ resolution, SIRAS), four TI⁺ data sets (soaked at 2 mM and 20 mM concentrations, diffraction to 3.8–4.1 $\text{\AA}$, MAD), one data set of the selenomethionine-substituted U1A protein (diffraction to 3.5 $\text{\AA}$, MAD), one data set of a crystal grown with an oligonucleotide containing a bromine in place of the dT-1 methyl group (diffraction to 3.7 $\text{\AA}$, MAD), one methyl mercury data set (soaked at 1 mM, diffraction to 3.7 $\text{\AA}$, MAD), and two Ta₆Br₁₄ data sets (soaked at 2 mM, diffraction to 3.65–3.7 $\text{\AA}$, SAD).

All data were post-scaled using the local-scaling routine NEWLSC (M. Rould, A. M. Friedman, J.W., and T. A. Steitz, unpublished results). Initial experimental phases were determined using the Yb³⁺ SIRAS, two TI⁺ MAD, and one TI⁺ SAD data sets. The phases were improved by including all derivatives in heavy-atom parameter refinement using MLPHARE⁴⁷ with 'external' phases obtained from density modification⁴⁸. All phase probabilities were combined using MLPHARE and SIGMAA of CCP4⁴⁷. Density modification protocols were undertaken using DM and DMMULTI of CCP4⁴⁷ and CNS⁴⁵. During the phase-improvement process, we observed that amplitude sharpening, applied with a negative B -factor value of $-50 \text{ \AA}^2$ to the observed data, substantially improved the quality of the resulting experimental maps. The CNS density modification routine successfully extended the experimental phases to 3.1- $\text{\AA}$ resolution. The quality of the resulting experimental electron-density maps was sufficient to interpret the entire structure, and the model was built using the program O⁴⁹. Strong density for the phosphoribose backbone was visible throughout the molecule and base density was visible for the vast majority of nucleotides (Fig. 3a).

The structure refinement was carried out using the program CNS⁴⁵. Metal ions in the structure were identified on the basis of relevant heavy-atom binding sites (Tb³⁺, Yb³⁺, Mn²⁺ and TI⁺ soaks), temperature factors during refinement, and coordination geometry. All residues in the intron/exon complex and all but three residues of the U1A protein were sufficiently ordered for inclusion in the final model. Two regions had substantially higher average B factors than the rest of the molecule. These included the P5 G71–C74 tetraloop and the majority of the P6a helix below the U101–U118 mispair, including the U1A protein. These two regions of the molecule pack against each other within the lattice. The region with the lowest B factors included the nucleotides surrounding the active site.

Received 26 February; accepted 12 May 2004; doi:10.1038/nature02642.
Published online 2 June 2004.

- Cech, T. R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44**, 207–210 (1986).
- Cech, T. R., Zaugg, A. J. & Grabowski, P. J. *In vitro* splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496 (1981).
- Cech, T. R. & Golden, B. L. In *The RNA World* 2nd edn (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 321–349 (Cold Spring Harbor Laboratory, New York, 1999).
- Reinhold-Hurek, B. & Shub, D. A. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**, 173–176 (1992).
- Cate, J. H. *et al.* Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
- Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282**, 259–264 (1998).
- Bass, B. L. & Cech, T. R. Ribozyme inhibitors: deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the *Tetrahymena* ribosomal ribonucleic acid precursor. *Biochemistry* **25**, 4473–4477 (1986).
- Moran, S., Kierzek, R. & Turner, D. H. Binding of guanosine and 3' splice site analogues to a group I ribozyme: interactions with functional groups of guanosine and with additional nucleotides. *Biochemistry* **32**, 5247–5256 (1993).
- Nagai, K., Oubridge, C., Jessen, T. H., Li, J. & Evans, P. R. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515–520 (1990).
- Ferre-D'Amare, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556 (2000).
- Tanner, M. A. & Cech, T. R. Activity and thermostability of the small self-splicing group I intron in the pre-tRNA^{leu} of the purple bacterium *Azoarcus*. *RNA* **2**, 74–83 (1996).
- Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610 (1990).
- Rangan, P., Masquida, B., Westhof, E. & Woodson, S. A. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl Acad. Sci. USA* **100**, 1574–1579 (2003).
- McSwiggan, J. A. & Cech, T. R. Stereochemistry of RNA cleavage by the *Tetrahymena* ribozyme and evidence that the chemical step is not rate-limiting. *Science* **244**, 679–683 (1989).
- Nissen, P., Hansen, J., Ban, N., Moore, P. & Steitz, T. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Rupert, P. B. & Ferre-D'Amare, A. R. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, 780–786 (2001).
- Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342**, 391–395 (1989).
- Damberg, S. H. & Gutell, R. R. A comparative database of group I intron structures. *Nucleic Acids Res.* **22**, 3508–3510 (1994).
- Duarte, C. M., Wadley, L. M. & Pyle, A. M. RNA structure comparison, motif search and discovery

- using a reduced representation of RNA conformational space. *Nucleic Acids Res.* **31**, 4755–4761 (2003).
- Kuo, L. Y., Davidson, L. A. & Pico, S. Characterization of the *Azoarcus* ribozyme: tight binding to guanosine and substrate by an unusually small group I ribozyme. *Biochim. Biophys. Acta* **1489**, 281–292 (1999).
 - Strauss-Soukup, J. & Strobel, S. A. A chemical phylogeny of group I introns based upon interference mapping of a bacterial ribozyme. *J. Mol. Biol.* **302**, 339–358 (2000).
 - Soukup, J. K., Minakawa, N., Matsuda, A. & Strobel, S. A. Identification of A-minor tertiary interactions within a bacterial group I intron active site by 3-deazaadenosine interference mapping. *Biochemistry* **41**, 10426–10438 (2002).
 - Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H. & Moncoeur, E. Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nature Struct. Biol.* **5**, 60–66 (1998).
 - Herschlag, D., Eckstein, F. & Cech, T. R. The importance of being ribose at the cleavage site in the *Tetrahymena* ribozyme reaction. *Biochemistry* **32**, 8312–8321 (1993).
 - Strobel, S. A. & Cech, T. R. Minor groove recognition of the conserved G-U pair at the *Tetrahymena* ribozyme reaction site. *Science* **267**, 675–679 (1995).
 - Strobel, S. A. & Ortoleva-Donnelly, L. A hydrogen bonding triad stabilizes the chemical transition state of a group I ribozyme. *Chem. Biol.* **6**, 153–156 (1999).
 - Ban, N., Nissen, P., Hansen, J., Moore, P. & Steitz, T. The complete atomic structure of the large ribosomal subunit at 2.4 $\text{\AA}$ resolution. *Science* **289**, 905–920 (2000).
 - Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
 - Shan, S., Yoshida, A., Sun, S., Piccirilli, J. A. & Herschlag, D. Three metal ions at the active site of the *Tetrahymena* group I ribozyme. *Proc. Natl Acad. Sci. USA* **96**, 12299–12304 (1999).
 - Shan, S., Kravchuk, A. V., Piccirilli, J. A. & Herschlag, D. Defining the catalytic metal ion interactions in the *Tetrahymena* ribozyme reaction. *Biochemistry* **40**, 5161–5171 (2001).
 - Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* **362**, 85–88 (1993).
 - Weinstein, L. B., Jones, B. C. N. M., Cossick, R. & Cech, T. R. A second catalytic metal ion in a group I ribozyme. *Nature* **388**, 805–808 (1997).
 - Yoshida, A., Sun, S. & Piccirilli, J. A. A new metal ion interaction in the *Tetrahymena* ribozyme reaction revealed by double sulfur substitution. *Nature Struct. Biol.* **6**, 318–321 (1999).
 - Sjogren, A. S., Pettersson, E., Sjoberg, B. M. & Stromberg, R. Metal ion interaction with cosubstrate in self-splicing of group I introns. *Nucleic Acids Res.* **25**, 648–653 (1997).
 - Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90**, 6498–6502 (1993).
 - Szewczak, A. A., Kosek, A. B., Piccirilli, J. A. & Strobel, S. A. Identification of an active site ligand for a group I ribozyme catalytic metal ion. *Biochemistry* **41**, 2516–2525 (2002).
 - Basu, S. *et al.* A specific monovalent metal ion integral to the A-A platform of the RNA tetraloop receptor. *Nature Struct. Biol.* **5**, 986–992 (1998).
 - Rangan, P. & Woodson, S. A. Structural requirement for Mg²⁺ binding in the group I intron core. *J. Mol. Biol.* **329**, 229–238 (2003).
 - Grosshans, C. A. & Cech, T. R. Metal ion requirements for sequence-specific endoribonuclease activity of the *Tetrahymena* ribozyme. *Biochemistry* **28**, 6888–6894 (1989).
 - Knitt, D. S. & Herschlag, D. pH dependencies of the *Tetrahymena* ribozyme reveal an unconventional origin of an apparent pK_a. *Biochemistry* **35**, 1560–1570 (1996).
 - Shan, S. O. & Herschlag, D. Probing the role of metal ions in RNA catalysis: kinetic and thermodynamic characterization of a metal ion interaction with the 2'-moiety of the guanosine nucleophile in the *Tetrahymena* group I ribozyme. *Biochemistry* **38**, 10958–10975 (1999).
 - Steitz, T. A. A mechanism for all polymerases. *Nature* **391**, 231–232 (1998).
 - Sigel, R. K., Vaidya, A. & Pyle, A. M. Metal ion binding sites in a group II intron core. *Nature Struct. Biol.* **7**, 1111–1116 (2000).
 - Sontheimer, E. J., Sun, S. & Piccirilli, J. A. Metal ion catalysis during splicing of premessenger RNA. *Nature* **388**, 801–805 (1997).
 - Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
 - Otwinski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
 - Collaborative Computational Project 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 - Rould, M. A., Perona, J. J. & Steitz, T. A. Improving multiple isomorphous replacement phasing by heavy-atom refinement using solvent-flattened phases. *Acta Crystallogr. A* **48**, 751–756 (1992).
 - Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 - Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Steitz, T. Steitz, J. Piccirilli, A. Pyle, S. Woodson, L. Szewczak, J. Cochran, M. Gill, R. Anderson and A. Seila for discussion and comments on the manuscript; C. Höbartner and R. Micura for the gift of a 2'-seleno-methyl substituted dCIRC oligonucleotide (Supplementary Information); L. Wadley for assistance with PRIMOS analysis; M. Becker and the staff of the X-25 beamline at Brookhaven NSLS for assistance with data collection; and the staff in the Yale Center for Structural Biology for extensive technical assistance. This project was supported by grants from the NSF and the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A.S. (strobel@csb.yale.edu) or J.W. (wang@csb.yale.edu). Coordinates for the *Azoarcus* pre-2S complex have been deposited in the Protein Data Bank under accession code 1T42.

No stellar p-mode oscillations in space-based photometry of Procyon

Jaymie M. Matthews¹, Rainer Kusching¹, David B. Guenther², Gordon A. H. Walker¹, Anthony F.J. Moffat³, Slavek M. Rucinski⁴, Dimitar Sasselov⁵ & Werner W. Weiss⁶

¹Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver V6T 1Z1, Canada

²Institute for Computational Astrophysics, Department of Astronomy and Physics, St Mary's University, Halifax, Nova Scotia B3H 3C3, Canada

³Département de physique, Université de Montréal, CP 6128, Succ. Centre-Ville, Montréal H3C 3J7, Canada

⁴Department of Astronomy and Astrophysics, University of Toronto, 60 St George Street, Toronto M5S 3H8, Canada

⁵Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

⁶Institut für Astronomie, Universität Wien, Türkenschanzstrasse 17, A-1180 Wien, Austria

Pressure-driven (p-mode) oscillations at the surface of the Sun, resulting from sound waves travelling through the solar interior, are a powerful probe of solar structure, just as seismology can reveal details about the interior of the Earth. Astronomers have hoped to exploit p-mode asteroseismology¹ in Sun-like stars to test detailed models of stellar structure and evolution, but the observations are extremely difficult. The bright star Procyon has been considered one of the best candidates for asteroseismology, on the basis of models and previous reports^{2–8} of p-modes detected in ground-based spectroscopy. Here we present a search for p-modes in 32 days of nearly continuous photometric satellite-based observations of Procyon. If there are p-modes in Procyon, they must have lifetimes less than 2–3 days and/or peak amplitudes <15 parts per million, which defy expectations from the Sun's oscillations and previous theoretical predictions. Target selection for future planned asteroseismology space missions may need to be reconsidered, as will the theory of stellar oscillations.

The MOST (Microvariability and Oscillations of Stars) micro-satellite⁹ is Canada's first orbiting space telescope and its first scientific research satellite to be launched in over 30 years. It houses a telescope of aperture 15 cm feeding a CCD photometer through a broad-band optical filter (350–700 nm). From the vantage point of its polar Sun-synchronous orbit (altitude, 820 km), MOST can monitor stars in a band of sky about 54° wide for up to 8 weeks without interruption. Starlight from each Primary Science Target (brighter than V-band magnitude $V \approx 6$) is projected onto the CCD as a fixed pupil image of the telescope covering about 1,500 pixels for high stability. The satellite was launched on 30 June 2003, and commissioned in the latter half of 2003. MOST is at present the only instrument capable of reaching photometric precision of a few parts per million (p.p.m.) in stars other than the Sun.

Principal goals of the MOST mission include the detection of p-modes in Sun-like stars and detection of reflected light from exoplanets¹⁰. MOST's first Primary Science Target was Procyon, a visual binary system containing an F5 subgiant primary ($V = 0.38$) and a white dwarf secondary, 5' away and more than 10 mag fainter. (Here we refer to the primary Procyon A simply as Procyon.) Procyon occupies an interesting position in the standard diagram of luminosity versus surface temperature. In Fig. 1 we show that evolutionary tracks for three different models of solar composition all pass through the observed position of Procyon. All are viable models, yet each is in a distinct evolutionary phase with potentially measurable differences in the eigenspectra or oscillation driving and damping characteristics. It was hoped that observing p-mode frequencies would enable us to distinguish between the possibilities.

There have been at least seven independent reports of p-mode detections in Procyon since 1986, based on high-resolution ground-based spectroscopy of the star^{2–8}. The longest of these data sets⁸ spanned 13 nights, covering 131 h (or about 42%) of the total timespan. All these observers found excess power in the Fourier spectra of their data in a frequency range of about 0.5–1.5 mHz (periods from about 33 to 11 min), with peak radial velocity (RV) amplitudes around 1 m s^{-1} . Large-amplitude p-modes had also been predicted for Procyon by theorists¹¹ as early as 1983. More recent calculations of stochastic excitation of p-modes by turbulent convection in the star^{12,13} suggest velocity amplitudes of about 1.3 m s^{-1} and luminosity amplitudes of about 60–70 p.p.m. (ref. 12). Even the most conservative estimate based on observed RV amplitudes and theoretical light-to-RV amplitude ratios predicts a luminosity amplitude of 20 p.p.m. (ref. 14). Therefore, this star qualified as a prime target for the MOST mission, whose oscillation detection limit for Procyon would be a few p.p.m.

We present in Fig. 2a the MOST photometry of Procyon collected from 8 January to 9 February 2004. Exposures of 0.9 s each were obtained at an average rate of 6 times per minute with only 7 h of gaps during 32 days — an unprecedented duty cycle of 99%. The best ground-based global campaigns to monitor stars other than the Sun, such as the Whole Earth Telescope, have never achieved duty cycles of more than 35% in one month. The MOST photometry is contaminated by scattered Earthlight entering the instrument focal plane, but this contributes less than 1% of the stellar signal from Procyon. This stray light is modulated by the known orbital period of the satellite ($P_{\text{orb}} = 101.413 \text{ min}$) so it is easy to distinguish from other sources of variability. The data presented here have undergone only a preliminary reduction which is more than sufficient to support our conclusions.

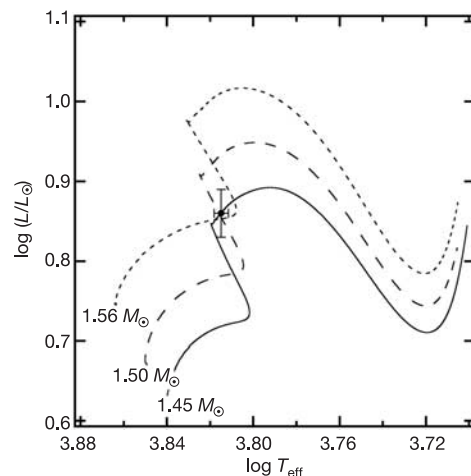


Figure 1 The location of Procyon in a plot of luminosity versus effective temperature. The basic physical parameters of Procyon are well specified: distance $d = 3.50 \pm 0.01 \text{ pc}$ based on a Hipparcos parallax, mass $M = 1.50 \pm 0.04 M_{\odot}$ from measurements of the binary orbit²¹, radius $R = 2.05 \pm 0.03 R_{\odot}$ determined by optical interferometry²², and temperature $T_{\text{eff}} = 6530 \pm 50 \text{ K}$ from model fits to its spectrum²³. ($M_{\odot}$, $R_{\odot}$ and $L_{\odot}$ are respectively the solar mass, radius and luminosity.) Three different stellar evolutionary tracks²⁴, all with primordial compositions of $(X, Z) = (0.71, 0.02)$ (where X and Z are the mass fractions of hydrogen and all elements heavier than helium, respectively) all pass through the observed²⁵ position of Procyon in the luminosity–temperature plane. The models are part of an extensive grid of models being computed to study the oscillation properties of stars²⁶. The $1.45 M_{\odot}$ model is $2.47 \times 10^9 \text{ yr} = 2.47 \text{ gigayear (Gyr)}$ old and has exhausted its core of hydrogen. The $1.50 M_{\odot}$ model is 2.20 Gyr old, and still has a mass fraction of hydrogen of $X_{\text{core}} = 0.005$ in its core. The $1.56 M_{\odot}$ model is 1.74 Gyr old, with $X_{\text{core}} = 0.138$. The depth of the convective envelope in Procyon severely tests the model physics because it is extremely shallow.

Figure 2b presents the MOST data binned into averages of the orbital period to better show the variability of the star on timescales of hours and days. To gauge instrumental drifts, we show in Fig. 2c the light curve of a fainter Secondary Science Target (HD 61199, $V = 8.2$) in the Procyon field, binned in the same way. We have discovered this star to be a multiperiodic variable of the δ Scuti type, whose main pulsation period is 57 min, with an amplitude of only 1.5 millimagnitudes (1,500 p.p.m.). It also exhibits a 3.9-day periodicity which we attribute to rotation or binarity. The slow trend common to Fig. 2b and c in the last third of the data has been removed in the subsequent analysis. The variations seen in Fig. 2b at timescales less than a few days appear intrinsic to Procyon. A shorter segment of the Procyon light curve spanning only 0.3 day is plotted in Fig. 2d, to indicate the point-to-point scatter.

Figure 3a is the Fourier amplitude spectrum of the entire unbinned data set. The peaks at very low frequencies are due to the variations seen in Fig. 2b (except for the trend which has been removed). The notches in the spectrum correspond to the frequencies of the orbit and its harmonics, where power due to modulated stray light has been removed. Figure 3a does not reveal any

significant peaks with the characteristic roughly equal spacings of p-modes¹. Our detection thresholds for coherent oscillations are about 3 p.p.m. at frequencies above 2 mHz, and 7 p.p.m. between 0.5 and 2.0 mHz. This is the best photometric precision so far reported in the literature^{15–17} by about an order of magnitude.

In the Sun, individual p-mode oscillations have finite lifetimes typically of days to weeks¹⁸ and do not remain coherent. However, in extended photometry of the Sun from space¹⁹, a characteristic p-mode pattern of roughly equal spacing clearly rises above the noise with amplitudes up to ~ 6 p.p.m. To test how sensitive is the complete MOST data set to oscillations of finite lifetimes, we embedded in the raw data modulated sine waves of known frequencies, whose amplitudes were allowed to grow and decay with a range of lifetimes, and whose phases were random. We then processed these data in the same way as before. As expected, oscillations which have larger amplitudes for longer times are more easily recovered in the Fourier spectrum of the data (see Supplementary Information). We show in Fig. 3b a conservative case of modes with lifetimes of only 3 days and maximum input amplitudes between 15 and 30 p.p.m., where the frequencies are those reported in the most extensive RV studies^{6,8}. The presence of p-modes would be recognisable in our data even by eye for modes with lifetimes as short as this.

Perhaps the average mode lifetimes in Procyon are indeed much shorter than in the Sun? To explore this, we divided the data into various subsets to look for recurring frequencies and/or an overall pattern formed by intermittent frequencies seen in short data segments. We repeated the experiments with simulations (sampled identically to the MOST data) containing pink noise, whose amplitude scales inversely with frequency. This is roughly what is expected in the light variations due to granulation, the pattern of

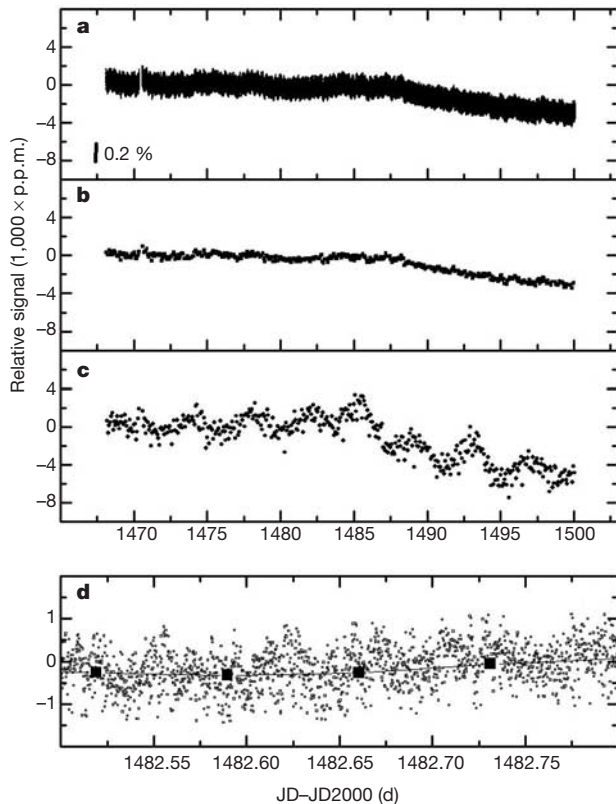


Figure 2 The light curve of Procyon. **a**, All 231,524 0.9-s exposures spanning 32 days. For the first 21 days, exposures were obtained 4 times per minute; for the remaining time, at twice this rate. The mean signal level per exposure is 1.34×10^7 ADU (analogue-to-digital units). The gaps in this light curve total only 7 hours owing to brief interruptions of satellite pointing. **b**, The same data, binned at intervals of the satellite orbital period of 101.413 min. We estimate the r.m.s. point-to-point uncertainty in the unbinned data to be about 500 p.p.m.; in the binned data, 75 p.p.m., so most of the structure seen in Fig. 1b is signal, not noise. (We are not yet at the limit of Poisson statistics, for which the r.m.s. values are about 275 and 11 p.p.m., respectively.) **c**, The light curve of a fainter star in the Procyon field, binned in the same way as **b**. These data show an intrinsic 3.9-day period visible in the light curve, and a main pulsation period of 57 min, which is responsible for much of the apparent scatter. **d**, An expanded view of 0.3 day of the data in Fig. 3a and b. The small open squares are the unbinned photometry; the large filled squares connected by straight lines are the data binned at the orbital period. JD, Julian day.

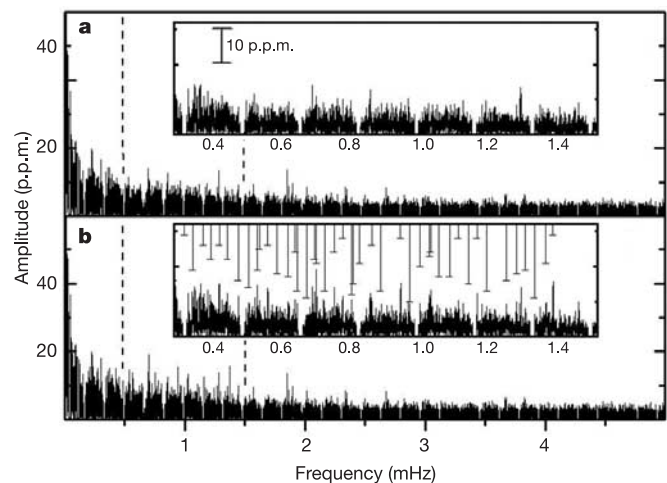


Figure 3 The Fourier amplitude spectrum of Procyon and the detection sensitivity of the MOST data. **a**, The spectrum of the photometry presented in Fig. 2a; the maximum frequency corresponds to a period of 2.8 min. (There is no excess power above the noise out to the Nyquist frequency of about 50 mHz.) The dashed vertical lines mark the range in which p-modes were reported by previous observers. The inset is an expanded view of that frequency range. The notches in the noise level of the spectra are where power has been filtered at the satellite orbital frequency and its harmonics to remove known stray light artefacts. **b**, The spectrum of the photometry in which artificial signals were embedded prior to data processing. The signals simulate stochastically excited p-modes with a lifetime of only 3 days and peak amplitudes between 15 and 30 p.p.m. The frequencies embedded are from previously reported detections^{6,8}. The inset is at the same scale as in **a**. The short horizontal bars represent the input amplitudes, with vertical lines extending above to make the input frequency spacings more obvious. Enough of the input mode frequencies – even with both lifetimes and amplitudes lower than expected from theory – are recovered to distinguish the simulation from the real data in **a**.

brighter rising turbulent cells of gas due to surface convection in Sun-like stars. We also applied the same 'comb response' test to our data as performed by Mosser *et al.*⁵ (see their equation (4)), searching for a comb-like pattern of equal spacing around the largest peaks in the power spectrum in the frequency range of interest. None of these tests reveals compelling evidence of anything other than granulation noise in our data (see Supplementary Information).

To make a more direct comparison between the MOST photometry and previous RV measurements of Procyon, we have sampled and filtered the MOST data in the same way as the ground-based data and compared the resulting Fourier power spectra to plots in the literature. Figure 4a is the power spectrum of RV measurements of Procyon published by Martić *et al.*⁶ which was used as evidence for p-modes. Other data samples and power spectra^{2–8} and their treatments are very similar in character. Figure 4b is the power spectrum of the MOST photometry of Procyon sampled at the same restricted time intervals as the Martić *et al.* spectroscopy, and filtered in the same way. Both Fig. 3a and b show a similar apparent excess in power between 0.5 and 1.5 mHz; the same experiment applied to other segments of the MOST data yield similar results. We suggest the possibility that apparent p-mode power identified by ground-based observers may be an artefact of insufficient gapped sampling of granulation variations.

There is certainly no convincing evidence of p-modes in the MOST data for Procyon, whose duty cycle of 99% for 32 days makes it the most sensitive data set for asteroseismology available for any star other than the Sun. Any oscillations present must have lifetimes less than 2–3 days and/or peak amplitudes below about 15 p.p.m. This null photometric result has several implications. (1) Models of stochastic excitation of acoustic oscillations in stars will be strongly constrained by the surprisingly short lifetimes of the p-modes in Procyon or by their very low amplitudes (or absence altogether). (2) The previous detections of p-modes in Procyon should be re-examined in the context of the new photometry. If those RV detections prove to be genuine, the MOST results eliminate several published theoretical^{12,13} p-mode amplitude relations and set tight limits on others¹⁴. If the RV detections are false, then the MOST data

supply the first measurements of the relative sensitivity of photometric and velocity data to granulation noise for a star other than the Sun. (3) In either case, the MOST results have a direct bearing on target selection as well as observing and analysis strategies for upcoming asteroseismology space missions like COROT²⁰, because those mission plans were based on pre-MOST expectations for Procyon that are no longer valid. □

Received 20 February; accepted 17 May 2004; doi:10.1038/nature02671.

1. Brown, T. M. & Gilliland, R. L. Asteroseismology. *Annu. Rev. Astron. Astrophys.* **32**, 37–82 (1994).
2. Walker, G. A. H. *et al.* The MOST asteroseismology mission: Ultraprecise photometry from space. *Publ. Astron. Soc. Pacif.* **115**, 1023–1035 (2003).
3. Green, D., Matthews, J. M., Seager, S. & Kuschnig, R. Scattered light from close-in extrasolar planets: Prospects of detection with the MOST satellite. *Astrophys. J.* **597**, 590–601 (2003).
4. Girard, T. M. *et al.* A redetermination of the mass of Procyon. *Astrophys. J.* **119**, 2428–2436 (2000).
5. Kervella, P. *et al.* The diameter and evolutionary state of Procyon A. *Astron. Astrophys.* **413**, 251–256 (2004).
6. Allende Prieto, C., Asplund, M., García López, R. J. & Lambert, D. L. Signatures of convection in the spectrum of Procyon: Fundamental parameters and iron abundance. *Astrophys. J.* **567**, 544–565 (2002).
7. Chaboyer, B., Demarque, P. & Guenther, D. B. The pulsation properties of Procyon A. *Astrophys. J.* **525**, 41–44 (1999).
8. Guenther, D. B. *et al.* Standard solar model. *Astrophys. J.* **387**, 372–393 (1992).
9. Guenther, D. B. & Brown, K. I. T. Matching stellar models to oscillation data. *Astrophys. J.* **600**, 419–434 (2004).
10. Gelly, B., Grec, G. & Fossat, E. Evidence for global pressure oscillations on Procyon. *Astron. Astrophys.* **164**, 383–394 (1986).
11. Brown, T. M. *et al.* Detection of possible p-mode oscillations on Procyon. *Astrophys. J.* **368**, 599–609 (1991).
12. Innis, J. L. *et al.* High-precision velocity observations of Procyon A – I. Search for p-mode oscillations from 1988, 1989 and 1990 observations. *Mon. Not. R. Astron. Soc.* **249**, 643–653 (1991).
13. Mosser, B., Millard, J. P., Mékarnia, D. & Gay, J. New limit on the p-mode oscillations of Procyon obtained by Fourier transform seismometry. *Astron. Astrophys.* **340**, 457–462 (1998).
14. Martić, M. *et al.* Evidence for global pressure oscillations on Procyon. *Astron. Astrophys.* **351**, 993–1002 (1999).
15. Carrier, F., Bouchy, F., Kienle, F. & Blecha, A. in *Radial and Nonradial Pulsations as Probes of Stellar Physics* (eds Aerts, C., Bedding, T. R. & Christensen-Dalsgaard, J.) 468–469 (Astronomical Society of the Pacific, San Francisco, 2002).
16. Martić, M., Lebrun, J.-C., Appourchaux, T. & Korzennik, S. G. p-mode frequencies in solar-like stars: I. Procyon A. *Astron. Astrophys.* (in the press).
17. Christensen-Dalsgaard, J. & Frandsen, S. Stellar 5-minute oscillations. *Sol. Phys.* **82**, 469–486 (1983).
18. Houdek, G., Balmforth, N. J., Christensen-Dalsgaard, J. & Gough, D. O. Amplitudes of stochastically excited oscillations in main-sequence stars. *Astron. Astrophys.* **351**, 582–596 (1999).
19. Samadi, R., *et al.* in *First Eddington Workshop on Stellar Structure and Habitable Planet Finding* (eds Favata, F., Roxburgh, I. W. & Galadi-Enríques, D.) 87–94 (ESA SP-485, ESA Publications Division, Noordwijk, 2002).
20. Kjeldsen, H. & Bedding, T. R. Amplitudes of stellar oscillations: The implications for asteroseismology. *Astron. Astrophys.* **293**, 87–106 (1995).
21. Gilliland, R. L. *et al.* A search for solar-like oscillations in the stars of M67 with CCD ensemble photometry on a network of 4m telescopes. *Astron. J.* **106**, 2441–2476 (1993).
22. Kurtz, D. W. Frequency variability in the rapidly oscillating Ap star HR 3831: three more years of monitoring. *Mon. Not. R. Astron. Soc.* **287**, 69–78 (1997).
23. Retter, A., Bedding, T. R., Buzasi, D. L., Kjeldsen, H. & Kiss, L. L. A tight upper limit on oscillations in the Ap star ϵ Ursae Majoris from WIRE photometry. *Astrophys. J.* **601**, L95–L98 (2004).
24. Toutain, T. & Fröhlich, C. Characteristics of solar p-modes: results from the IPIR experiment. *Astron. Astrophys.* **257**, 287–297 (1992).
25. Fröhlich, C. *et al.* First results from VIRGO, the experiment for helioseismology and solar irradiative monitoring on SOHO. *Sol. Phys.* **170**, 1–25 (1997).
26. Baglin, A., *et al.* in *Proc. First Eddington Workshop on Stellar Structure and Habitable Planet Finding* (eds Favata, F., Roxburgh, I. W. & Galadi-Enríques, D.) 17–24 (ESA SP-485, ESA Publications Division, Noordwijk, 2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The MOST project is funded and supervised by the Canadian Space Agency. We are grateful to the engineers at Dynacon Inc., the University of Toronto Institute for Aerospace Studies – Space Flight Laboratory, the University of British Columbia Department of Physics and Astronomy, and the University of Vienna and Austrian Technical University (Vienna) for their contributions to the success of the MOST mission. J.M.M., D.B.G., G.A.H.W., A.F.J.M. and S.M.R. were supported by the Natural Sciences and Engineering Research Council Canada. W.W.W. received support from the Austrian Space Agency and the Science Fund. We are grateful for comments and suggestions from T. Brown and J.-P. Maillard.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.M. (matthews@astro.ubc.ca).

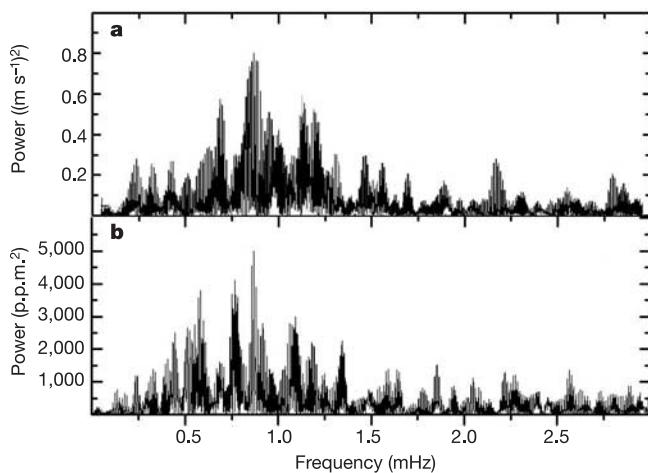


Figure 4 The effects of incomplete sampling on observations of Procyon. **a**, The power spectrum of ground-based radial velocity data published by Martić *et al.*⁴ (their Fig. 7, reproduced with permission). The data to produce this spectrum were obtained on 4 nights during 7–11 November 1998; a total of 24.9 hours over a span of 4 days (a duty cycle of about 26%). A high-pass filter was applied by those authors to their data to suppress power at low frequencies from stellar granulation and observational drifts. **b**, A typical power spectrum of MOST photometry sampled and filtered in the same way as the data in **a**.

Experimental demonstration of five-photon entanglement and open-destination teleportation

Zhi Zhao¹, Yu-Ao Chen¹, An-Ning Zhang¹, Tao Yang¹, Hans J. Briegel² & Jian-Wei Pan^{1,3}

¹Department of Modern Physics and Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui, 230026, China

²Institute for Theoretical Physics, University of Innsbruck, and Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, A-6020 Innsbruck, Austria

³Physikalisches Institut, Universität Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany

Quantum-mechanical entanglement of three^{1,2} or four^{3,4} particles has been achieved experimentally, and has been used to demonstrate the extreme contradiction between quantum mechanics and local realism^{5,6}. However, the realization of five-particle entanglement remains an experimental challenge. The ability to manipulate the entanglement of five or more particles is required^{7,8} for universal quantum error correction. Another key process in distributed quantum information processing^{9,10}, similar to encoding and decoding, is a teleportation protocol^{11,12} that we term ‘open-destination’ teleportation. An unknown quantum state of a single particle is teleported onto a superposition of N particles; at a later stage, this teleported state can be read out (for further applications) at any of the N particles, by a projection measurement on the remaining particles. Here we report a proof-of-principle demonstration of five-photon entanglement and open-destination teleportation (for $N = 3$). In the experiment, we use two entangled photon pairs to generate a four-photon entangled state, which is then combined with a single-photon state. Our experimental methods can be used for investigations of measurement-based quantum computation^{9,10} and multi-party quantum communication^{13,14}.

In our experiment, the way to entangle five photons is a straightforward generalization of the schemes suggested for observation of three- and four-photon Greenberger–Horne–Zeilinger (GHZ) entanglement^{15,16}. We start from two polarization-entangled photon pairs, 2–3 and 4–5, which are in the state $|\Phi^+\rangle$ (Fig. 1a). We use the usual Bell states

$$|\Phi^\pm\rangle_{ij} = \frac{1}{\sqrt{2}}(|H\rangle_i|H\rangle_j \pm |V\rangle_i|V\rangle_j) \\ |\Psi^\pm\rangle_{ij} = \frac{1}{\sqrt{2}}(|H\rangle_i|V\rangle_j \pm |V\rangle_i|H\rangle_j) \quad (1)$$

where H and V denote respectively horizontal and vertical linear polarizations, and i and j index the spatial modes of the photons. One photon out of each pair (3 and 4) is then steered to a polarizing beam splitter (PBS) where the path lengths of each photon have been adjusted such that they arrive simultaneously. Because the PBS transmits H and reflects V polarization, coincidence detection between the two outputs of PBS₃₄ implies that either both photons 3 and 4 are H -polarized or both are V -polarized, and thus projects the four-photon state onto a two-dimensional subspace spanned by $|H\rangle_2|H\rangle_3|H\rangle_4|H\rangle_5$ and $|V\rangle_2|V\rangle_3|V\rangle_4|V\rangle_5$. After PBS₃₄, the renormalized state corresponding to a fourfold coincidence (before photon 2 passes through PBS₁₂) is

$$|\Phi\rangle_{2345} = \frac{1}{\sqrt{2}}(|H\rangle_2|H\rangle_3|H\rangle_4|H\rangle_5 + |V\rangle_2|V\rangle_3|V\rangle_4|V\rangle_5) \quad (2)$$

which exhibits four-photon GHZ entanglement^{4,15}.

To generate five-photon GHZ entanglement, we further prepare

photon 1 in the state $\frac{1}{\sqrt{2}}(|H\rangle_1 + |V\rangle_1)$ and adjust the path lengths of photons 1 and 2 such that they also arrive at PBS₁₂ simultaneously. Then, we see that after the photons pass through the two PBS, the state corresponding to a fivefold coincidence is given by:

$$|\Phi\rangle_{12345} = \frac{1}{\sqrt{2}}(|H\rangle_1|H\rangle_2|H\rangle_3|H\rangle_4|H\rangle_5 \\ + |V\rangle_1|V\rangle_2|V\rangle_3|V\rangle_4|V\rangle_5) \quad (3)$$

This is exactly a five-photon GHZ state.

The scheme described above deserves some further comments. First, the five-photon entanglement here is observed only under the condition that there is one and only one photon in each of the five output modes. This post-selective feature, however, does not prevent us from performing an experimental test of quantum non-locality^{5,6} or in-principle verification of elements of linear optical quantum information processing^{17,18}.

Moreover, the set-up sketched in Fig. 1a can also be used to realize open-destination teleportation, which can be viewed as encoding followed by decoding. In the open-destination teleportation scheme, we use the four-particle GHZ state (equation (2)) as the resource and photon 1 as the ‘teleportee’. To see this, we consider a specific example (Fig. 1b). Suppose that photon 1 is in a general unknown polarization state $|\Psi\rangle_1 = \alpha|H\rangle_1 + \beta|V\rangle_1$. In order to overcome the unavoidable decoherence effect on the state $|\Psi\rangle_1$, one could exploit quantum error correction. For example, to detect and correct the bit-flip error^{8,19}, we could encode the unknown state of photon 1 into a three-qubit superposition, say $|\Psi\rangle_{345} = \alpha|H\rangle_3|H\rangle_4|H\rangle_5 + \beta|V\rangle_3|V\rangle_4|V\rangle_5$. Note that, to realize universal

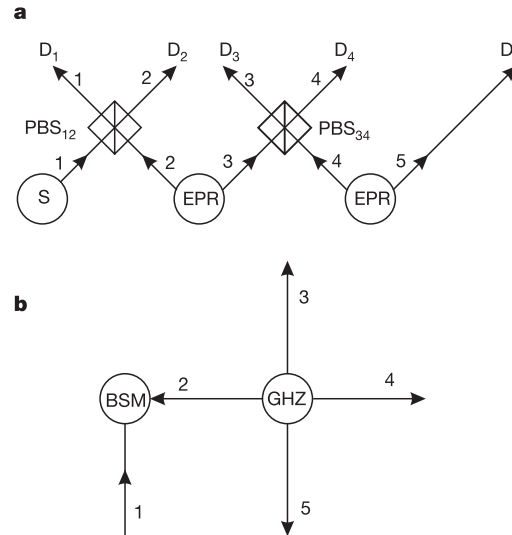


Figure 1 Diagrams showing the principles of generating five-photon entanglement and of achieving open-destination teleportation. **a**, Two Einstein-Podolsky-Rosen (EPR) entangled photon sources each emit one polarization-entangled photon pair, which is in the state $|\Phi^+\rangle = 1/\sqrt{2}(|H\rangle|H\rangle + |V\rangle|V\rangle)$. The single-photon source (S) emits a single photon state $\frac{1}{\sqrt{2}}(|H\rangle_1 + |V\rangle_1)$. After the photons 1, 2, 3 and 4 pass through the two polarizing beam splitters PBS₁₂ and PBS₃₄, conditional on detecting one photon in each of the five output modes, the five photons will exhibit five-photon Greenberger-Horne-Zeilinger (GHZ) entanglement. **b**, The GHZ entanglement source emits a four-photon maximally entangled state $|\Phi\rangle_{2345} = \frac{1}{\sqrt{2}}(|H\rangle_2|H\rangle_3|H\rangle_4|H\rangle_5 + |V\rangle_2|V\rangle_3|V\rangle_4|V\rangle_5)$. Photon 1 is in a general unknown polarization state $|\Psi\rangle_1 = \alpha|H\rangle_1 + \beta|V\rangle_1$. By performing a joint Bell-state measurement (BSM) between photons 1 and 2, we can follow the teleportation protocol to transfer the initial state of photon 1 to a multi-particle superposition $|\Psi\rangle_{345} = \alpha|H\rangle_3|H\rangle_4|H\rangle_5 + \beta|V\rangle_3|V\rangle_4|V\rangle_5$. Moreover, by further performing a polarization analysis on any two of the three photons 3, 4 and 5 in the $\pm$ basis one can convert the remaining photon into the initial state of photon 1.

quantum error correction, a more complicated encoding procedure is needed.

To achieve the three-qubit encoding, we first prepare photons 2, 3, 4 and 5 in the four-photon GHZ state $|\Phi\rangle_{2345}$ of equation (2). Then, using the four Bell states of photons 1 and 2 the overall state of the five photons can be rewritten as:

$$\begin{aligned} |\Psi\rangle_{12345} &= |\Psi\rangle_1 |\Phi\rangle_{2345} \\ &= \frac{1}{2} [|\Phi^+\rangle_{12} (\alpha |H\rangle_3 |H\rangle_4 |H\rangle_5 + \beta |V\rangle_3 |V\rangle_4 |V\rangle_5) \\ &\quad + |\Phi^-\rangle_{12} (\alpha |H\rangle_3 |H\rangle_4 |H\rangle_5 - \beta |V\rangle_3 |V\rangle_4 |V\rangle_5) \\ &\quad + |\Psi^+\rangle_{12} (\alpha |V\rangle_3 |V\rangle_4 |V\rangle_5 + \beta |H\rangle_3 |H\rangle_4 |H\rangle_5) \\ &\quad + |\Psi^-\rangle_{12} (\alpha |V\rangle_3 |V\rangle_4 |V\rangle_5 - \beta |H\rangle_3 |H\rangle_4 |H\rangle_5)] \end{aligned} \quad (4)$$

This implies that a joint Bell measurement on photons 1 and 2 would thus project the state of photons 3, 4 and 5 into one of the

four corresponding states, as shown in equation (4). Depending on the measurement results on photons 1 and 2, one can then perform a local unitary transformation, independent of $|\Psi\rangle_1$, on photons 3, 4 and 5 to convert the state into a three-particle superposition of $|\Psi\rangle_{345}$. This is how encoding works. During the encoding operation, the unknown state of a single particle is teleported onto a three-particle superposition, and can thus be protected against bit-flip error^{8,19}.

To demonstrate the working principle of the encoding operation, it is sufficient to identify one of the four Bell states (say, $|\Phi^+\rangle_{12}$) although this results in a reduced efficiency—the fraction of success—of 25%. The encoding operation can be achieved using the set-up shown in Fig. 1a. First, the set-up provides the necessary four-photon GHZ entanglement $|\Phi\rangle_{2345}$ after photons 3 and 4 passes through PBS₃₄. Second, as demonstrated in a recent experiment⁴, the projection onto state $|\Phi^+\rangle_{12}$ can be accomplished by performing a joint polarization measurement in the $+/-$ basis behind PBS₁₂, where $|\pm\rangle = \frac{1}{\sqrt{2}}(|H\rangle \pm |V\rangle)$. As stated in ref. 6, registering a $|+\rangle_1 |+\rangle_2$ or $|-\rangle_1 |-\rangle_2$ coincidence acts as a projection onto $|\Phi^+\rangle_{12}$. This projection leaves photons 3, 4 and 5 in the state: $|\Psi\rangle_{345} = \alpha |H\rangle_3 |H\rangle_4 |H\rangle_5 + \beta |V\rangle_3 |V\rangle_4 |V\rangle_5$.

At a later stage, we can then use the decoding operation to read out or apply desired specific operations on the unknown state for further quantum information processing. To do so, we perform a local polarization measurement on two of the three photons 3, 4 and 5 in the $+/-$ basis. For example, if we perform a $\pm$ polarization analysis on photons 4 and 5, and if the measurement results on these two photons are the same—that is, $|+\rangle_4 |+\rangle_5$ or $|-\rangle_4 |-\rangle_5$ —then photon 3 is left in the state $|\Psi\rangle_3 = \alpha |H\rangle_3 + \beta |V\rangle_3$. If the results are opposite, namely $|+\rangle_4 |-\rangle_5$ or $|-\rangle_4 |+\rangle_5$, then photon 3 is left in the state $\alpha |H\rangle_3 - \beta |V\rangle_3$. In the second case, one can simply perform a local phase flip operation on photon 3 to convert its state into $|\Psi\rangle_3$, that is, the original state of photon 1. We can thus teleport the unknown quantum state from photon 1 to photon 3.

We note that, in a similar manner, the initial state of photon 1 can also be teleported either onto photon 4 or photon 5 by performing a polarization measurement either on photons 3 and 5 or on photons 3 and 4 in the $+/-$ basis. In contrast to the original teleportation scheme¹¹, after the encoding operation the destination of teleportation is left open until we perform a polarization measurement on two of the remaining three photons. This implies that, even though photons 3, 4 and 5 are far apart, one can still choose which particle should act as the output—that is, the particle to which the initial state of photon 1 is transferred¹². This is why we have called such an encoding-decoding procedure ‘open-destination’ teleportation. It is therefore a generalization of standard teleportation, in which no prior agreement on the final destination of the teleportation is necessary. Open-destination teleportation is possible if the parties initially share a multi-party entangled state, such as the GHZ state, but more generally any graph state²⁰ (of which the cluster state²¹ is another example) can be used. Graph states are natural generalizations of the Einstein-Podolsky-Rosen (EPR) state, which plays a key role in novel schemes of quantum information processing²¹. A potential application of the open-destination teleportation is the scheme of quantum computation introduced by Gottesman and Chuang⁹, using GHZ states and teleportation as a resource to realize quantum gates. Similarly, the ability to establish flexible teleportation channels with an entangled (cluster) state is a central ingredient in the one-way quantum computer¹⁰.

Although our scheme to manipulate five-photon entanglement is theoretically simple, its experimental realization is very challenging. So far, spontaneous parametric down-conversion (SPDC) is still the best source of entangled photons, which is therefore used as the basic entanglement resource in the present experiment. However, owing to the probabilistic feature of SPDC, the coincidence count rate in our five-photon experiment would be very low. To overcome

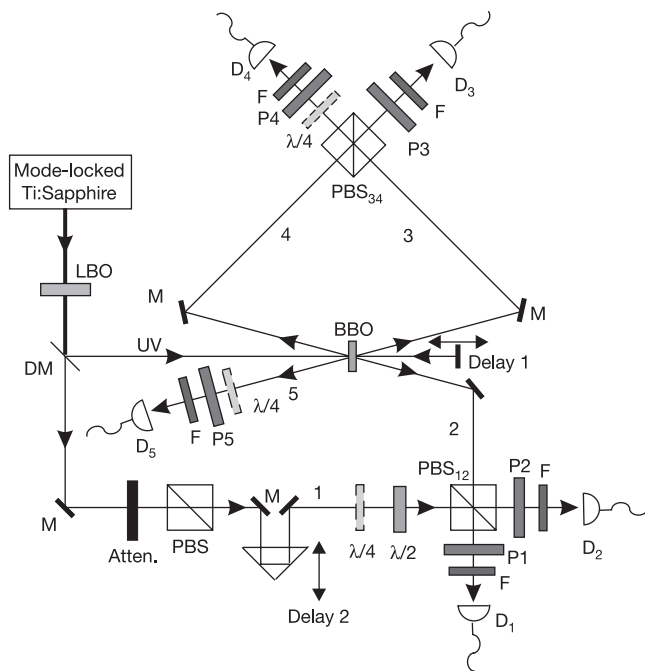


Figure 2 Set-up for experimental demonstration of five-photon entanglement and open-destination teleportation. To prepare the desired stable high-intensity source of entangled photons, various methods have been used. First, we properly focus the infrared pulse on the LBO crystal to achieve the best up-conversion efficiency. Meanwhile, to avoid damage to the LBO crystal caused by the focusing laser beam, we assemble the LBO crystal in a closed but transparent tube of oxygen. Second, by using a compact set-up and by focusing the UV pump onto the BBO crystal, we achieve both better collection efficiency and production rate of entangled photon pairs. Third, two CCD cameras (not shown) are used to monitor the laser beam direction. Thus, with the help of a feedback loop we can well stabilize the beam direction. In this way, we managed to obtain an average UV pump power of 480 mW and observe an average twofold coincidence of $2.4 \times 10^4 \text{ s}^{-1}$ both in modes 2–3 and 4–5. The mirrors Delay 1 and Delay 2 are used to adjust the time delay of photons 3 and 4 and of photons 1 and 2, respectively. In order to interfere photons 1 and 2 at PBS₁₂ and photons 3 and 4 at PBS₃₄, narrow band-width filters F ($\Delta\lambda_{\text{FWHM}} = 3 \text{ nm}$) and fibre-coupled detectors D1, ..., D5 have been used to ensure good temporal and spatial overlap of photons 1 and 2 and of photons 3 and 4. The wave plates $\lambda/2$ and $\lambda/4$ in front of PBS₁₂ are used to transform the horizontal polarization into linear $+/-$ or circular R/L polarization. The polarizers P1, ..., P5 oriented at $+/-$ basis and the $\lambda/4$ plate in front of the detectors allow measurement of linear $+/-$ or circular R/L polarization.

this difficulty, we combine a high-intensity source of entangled photons and a source of pseudo-single photons so that we can collect enough experimental data to confirm the existence of five-photon entanglement within a reasonable time.

A schematic drawing of our experimental set-up is shown in Fig. 2. In the experiment, a light pulse from a mode-locked Ti:sapphire laser (with a duration of 200 fs, a repetition rate of 76 MHz and a central wavelength of 788 nm) first passes through a frequency doubler, that is, the LBO crystal (LiB_3O_5). Behind the LBO, three dichroic beamsplitters (DM) are used to separate the

mixed ultraviolet (UV) and infrared light components. The UV pulse further passes through a β -barium borate (BBO) crystal twice to generate two entangled photon pairs in the input modes 2–3 and 4–5²², where both pairs are in the desired Bell state, $|\Phi^+\rangle$. The

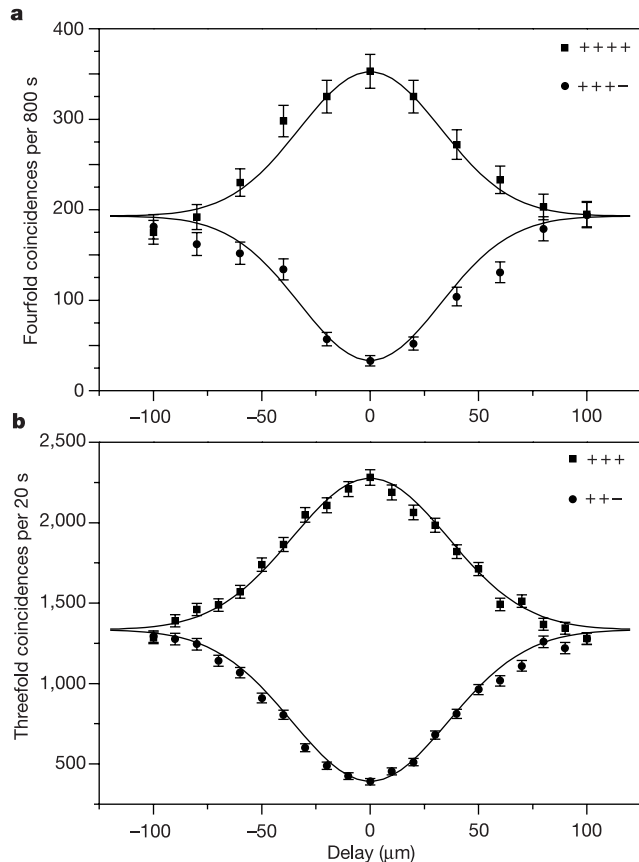


Figure 3 Experimental results showing the procedure used to achieve perfect temporal overlap for photons 1 and 2 and for photons 3 and 4. **a**, The two fourfold coincidence curves were obtained for the polarization settings of $|+\rangle_2|+\rangle_3|+\rangle_4|+\rangle_5$ and $|+\rangle_2|+\rangle_3|+\rangle_4|-\rangle_5$ as a function of the position of the Delay 1 mirror. During the measurement, PBS_{34} was inserted into the crossing position of modes 3 and 4 while PBS_{12} was lowered out of the crossing position of modes 1 and 2. Photon 2 was registered by D1. By fixing the Delay 1 mirror to the position where we observe the best four-photon interference visibility (that is, the centre of the peak and dip) we can achieve perfect temporal overlap for photons 3 and 4. Thus, the fourfold coincidence among the detectors D1, D3, D4 and D5 corresponds to the four-photon GHZ state of equation (2)¹⁵. The coherent superposition between the terms $|H\rangle_2|H\rangle_3|H\rangle_4|H\rangle_5$ and $|V\rangle_2|V\rangle_3|V\rangle_4|V\rangle_5$ was confirmed with an excellent interference visibility of 82%. **b**, The two threefold coincidence curves were obtained for the polarization settings of $|+\rangle_1|+\rangle_2|+\rangle_3$ and $|+\rangle_1|+\rangle_2|-\rangle_3$ as a function of the Delay 2 position. Here, PBS_{12} was inserted into the crossing position of modes 1 and 2 while PBS_{34} was lowered out of the crossing position of modes 3 and 4. Photon 3 was registered by D4. Similarly, by setting the Delay 2 position to the centre of the peak and dip of the threefold coincidence, we can achieve perfect temporal overlap for photons 1 and 2. At zero delay, the threefold coincidence among the detectors D1, D2 and D4 is a post-selective three-photon GHZ entanglement¹⁶. In our three-photon entanglement, the visibility was observed to be 68%. After achieving the perfect time overlap between photons 1 and 2, PBS_{34} was raised back to its original position to implement five-photon entanglement. The error bars here and in Fig. 4 are defined as the square root of the observed multi-fold coincidence.

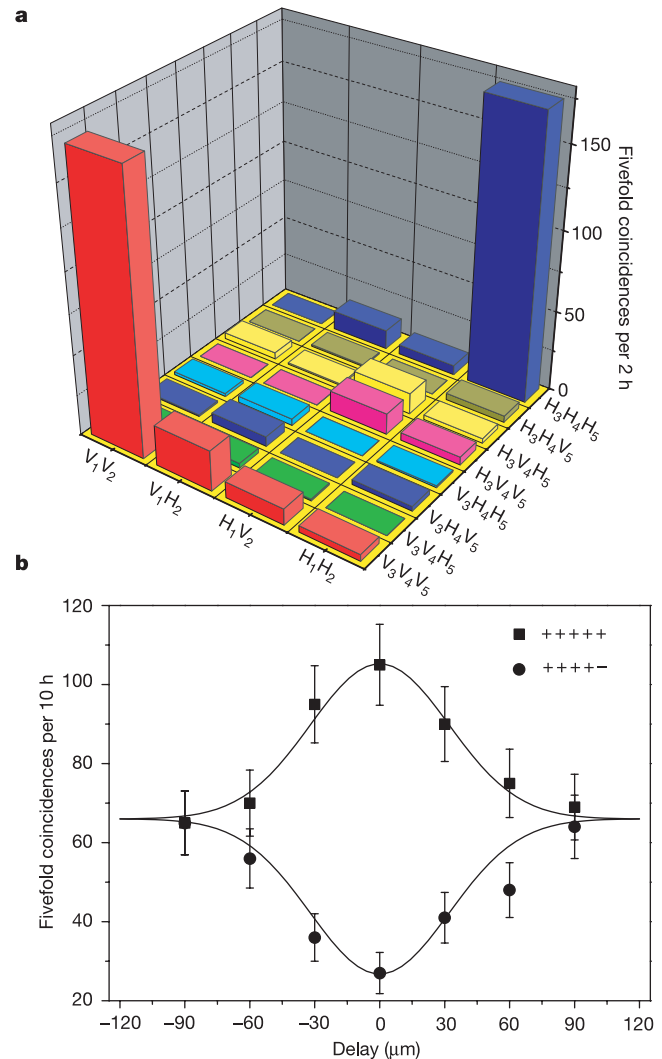


Figure 4 Experimental results for the observation of five-photon entanglement. **a**, To verify that five-photon GHZ entanglement has been successfully generated, we first measure the 32 possible components in the H/V basis. The measurement results show that the signal-to-noise ratio, that is, the ratio of any of the desired components to any of the 30 other non-desired ones, is about 40:1 on average. This confirms that within the experimental precision only the desired $HHHHH$ and $VVVVV$ terms are present. **b**, To further demonstrate that these two terms are indeed in a coherent superposition, we also perform a joint polarization measurement on all five photons in the $+/ -$ basis. The two curves were obtained for the polarization settings of $|+\rangle_1|+\rangle_2|+\rangle_3|+\rangle_4|+\rangle_5$ and $|+\rangle_1|+\rangle_2|+\rangle_3|+\rangle_4|-\rangle_5$ as a function of the position of the Delay 1 mirror with the position of the Delay 2 mirror fixed, which confirms the coherent superposition of the $HHHHH$ and $VVVVV$ terms with a visibility of 59%. In the experiment, the non-desired components are mainly due to the limited quality of the polarization optics and the double emissions in the source of single and entangled photons. The double emissions contribute around 3% to overall fivefold events. Without additional assumptions, it is not possible to use the above data to give an explicit estimation of the fidelity of our five-photon entanglement. However, if we are allowed to assume that the non-desired components in the H/V basis contribute equally to the $|+\rangle_1|+\rangle_2|+\rangle_3|+\rangle_4|+\rangle_5$ and $|+\rangle_1|+\rangle_2|+\rangle_3|+\rangle_4|-\rangle_5$ terms, the fidelity for our five-photon source to be in the state $|\Phi\rangle_{12345}$ is estimated to be 68%. Assuming that the non-desired components only contribute to the $|+\rangle_1|+\rangle_2|+\rangle_3|+\rangle_4|-\rangle_5$ term, a lower bound of the fidelity is estimated to be 55%.

transmitted near-infrared pulse is attenuated by combining a beam attenuator ('Atten.' in Fig. 2) and a polarizing beam splitter (PBS) to a weak coherent state such that there is only a very small probability of containing a single photon for each pulse (~ 0.05 in our experiment). We can thus prepare the required single photon state¹⁶. Note that such a pseudo-single photon source has been used in numerous experiments^{23,24}. Throughout the experiment, the coincidence time-window is set to be 4 ns, which ensures that accidental coincidence is negligible.

In order to prepare a stable high-intensity source of entangled photons, various efforts have been made (see Fig. 2 legend). Compared to the source of entangled photons developed in the recent experiments²⁵, our source is not only brighter, but, most significantly, it is also ultra-stable—it can be easily stabilized for a couple of days.

Implementation of five-photon entanglement requires that photons 1 and 2 (3 and 4) have good spatial and temporal overlap at PBS₁₂ (PBS₃₄). To achieve this, the two outputs of PBS₁₂ and PBS₃₄ are spectrally filtered ($\Delta\lambda_{\text{FWHM}} = 3$ nm) and monitored by fibre-coupled single-photon detectors²⁶. Moreover, perfect temporal overlap is accomplished by observing interference fringes of three- and four-photon entanglement in the $+/ -$ basis (see Fig. 3 legend). The experimental three- and four-photon interference visibilities were observed to be 68% and 82%, respectively. Remarkably, the observed coincidence rate of our three-photon entanglement source is about 500 s^{-1} , which is three orders of magnitude higher than in a recent experiment²⁷. After achieving perfect time overlap between photons 1 and 2 and between photons 3 and 4, according to equation (3) the fivefold coincidence among the detectors D1, D2, D3, D4 and D5 will then exhibit five-photon GHZ entanglement.

To verify experimentally that five-photon entanglement has been successfully obtained, we first show that under the condition of registering a fivefold coincidence only the $|H\rangle|H\rangle|H\rangle|H\rangle|H\rangle$ and $|V\rangle|V\rangle|V\rangle|V\rangle|V\rangle$ components are observed, but no others. This was done by comparing the counts of all 32 possible polarization combinations, $|H\rangle|H\rangle|H\rangle|H\rangle|H\rangle$, ..., $|V\rangle|V\rangle|V\rangle|V\rangle|V\rangle$. The measurement results (Fig. 4a) in the H/V basis show that the signal-to-noise ratio, defined as the ratio of any of the desired components to any of the 30 other non-desired ones, is about 40:1 on average.

Second, we further perform a polarization measurement in the $+/ -$ basis to verify that the two terms of $|H\rangle|H\rangle|H\rangle|H\rangle|H\rangle$ and $|V\rangle|V\rangle|V\rangle|V\rangle|V\rangle$ are indeed in a coherent superposition. Transforming $|\Phi\rangle_{12345}$ to the $+/ -$ linear polarization basis yields an expression containing 16 (out of 32 possible) terms, each with an odd number of $|+\rangle$ components. Combinations with even numbers of $|+\rangle$ components do not occur. As a test for coherence we can check the presence or absence of various components. In Fig. 4b we compare the $|+\rangle|+\rangle|+\rangle|+\rangle|+\rangle$ and $|+\rangle|+\rangle|+\rangle|+\rangle|-\rangle$ count rates as a function of the Delay 1 mirror position (see Fig. 2) while the Delay 2 is at zero delay. When the Delay 1 is also at zero delay, the unwanted terms were found to be suppressed with an average visibility of 0.59 ± 0.07 , which is sufficient to violate a five-particle Bell-type inequality imposed by local realism²⁸. Therefore, the measurement results in Fig. 4 clearly demonstrate the existence of five-particle entanglement.

We now show how the same set-up can be used to implement

open-destination teleportation. To demonstrate that our open-destination teleportation protocol works for a general unknown polarization state of photon 1, we choose to teleport $|+\rangle$ and $|-\rangle$ linear polarization states $\frac{1}{\sqrt{2}}(|H\rangle \pm |V\rangle)$ and right-hand ($|R\rangle$) and left-hand ($|L\rangle$) circular polarization states $\frac{1}{\sqrt{2}}(|H\rangle \pm i|V\rangle)$. In the experiment, we decided to analyse the Bell state $|\Phi^+\rangle_{12}$. The required projection onto $|\Phi^+\rangle_{12}$ was achieved by registering a $|+\rangle_1|+\rangle_2$ coincidence behind PBS₁₂. Then, as already discussed, conditional on a $|+\rangle|+\rangle$ coincidence detection in the output modes 3 and 4, photon 5 will be left in the initial state of photon 1, that is, the unknown state of photon 1 is teleported to photon 5. In order to demonstrate that the initial state of photon 1 can also be teleported to some other location (for example, location 4 of Fig. 1b), we can, in a similar manner, perform a polarization analysis on photons 3 and 5 in the $+/ -$ linear polarization basis. Then, conditional on a $|+\rangle|+\rangle$ coincidence detection in the output modes 3 and 5, photon 4 will be left in the initial state of photon 1.

In our experiment, the integration time for each teleportation measurement is about 10 hours. As shown in Fig. 4b, in every polarization analysis measurement the fivefold coincidence rate is about 100 for the maximum (desired) component and 20 for the minimum (non-desired) component per 10 hours. The experimental fidelities of teleportation from photon 1 to photon 5 and from photon 1 to photon 4 for $+/ -$ linear and R/L circular polarization states are shown in Table 1. From Table 1, one can see that all the observed teleportation fidelities ($\sim 0.80 \pm 0.04$) are well above the classical limit of two-thirds, hence fully demonstrating open-destination teleportation.

Although compared to the previous experiments^{1,4,29} our experimental demonstration of five-photon entanglement and open-destination teleportation might seem to be only a modest step forward, the implications are profound. First, our experiment has demonstrated the ability to manipulate five-particle entanglement, which is the threshold number of qubits required for universal error correction^{7,8}. Second, the realization of open-destination teleportation opens up new possibilities for distributed quantum information processing^{9,10}. Last, the techniques developed in the present experiment enable experimental investigations of a number of quantum protocols. Whereas our high-intensity three-photon entanglement source immediately allows an experimental realization of quantum secret sharing^{13,14}, the five-photon experimental set-up that we report here should also allow experimental demonstrations of both bit-flip error rejection for quantum communication¹⁹ and a non-destructive CNOT gate for quantum computation³⁰. □

Received 12 February; accepted 10 May 2004; doi:10.1038/nature02643.

1. Bouwmeester, D., Pan, J.-W., Daniell, M., Weinfurter, H. & Zeilinger, A. Observation of three-photon Greenberger-Horne-Zeilinger entanglement. *Phys. Rev. Lett.* **82**, 1345–1349 (1999).
2. Rauschenbeutel, A. *et al.* Step-by-step engineered multiparticle entanglement. *Science* **288**, 2024–2028 (2000).
3. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
4. Pan, J.-W., Daniell, M., Gasparoni, S., Weihs, G. & Zeilinger, A. Experimental demonstration of four-photon entanglement and high-fidelity teleportation. *Phys. Rev. Lett.* **86**, 4435–4439 (2001).
5. Pan, J.-W., Bouwmeester, D., Daniell, M., Weinfurter, H. & Zeilinger, A. Experimental test of quantum non-locality in three-photon Greenberger-Horne-Zeilinger entanglement. *Nature* **403**, 515–519 (2000).
6. Zhao, Z. *et al.* Experimental violation of local realism by four-photon Greenberger-Horne-Zeilinger entanglement. *Phys. Rev. Lett.* **91**, 180401 (2003).
7. Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
8. Laflamme, R., Miquel, C., Paz, J. P. & Zurek, W. H. Perfect quantum error correcting code. *Phys. Rev. Lett.* **77**, 198–201 (1996).
9. Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
10. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
11. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **83**, 3081–3084 (1993).
12. Karlsson, A. & Bourennane, M. Quantum teleportation using three-photon entanglement. *Phys. Rev. A* **58**, 4394–4400 (1998).
13. Hillery, M., Buzek, V. & Berthiaume, A. Quantum secret sharing. *Phys. Rev. A* **59**, 1829–1834 (1999).
14. Scarani, V. & Gisin, N. Quantum communication between N partners and Bell's inequalities. *Phys. Rev. Lett.* **87**, 117901 (2001).

Table 1 Fidelities of open-destination teleportation at two different locations

Polarization	Location 4	Fidelities	Location 5
$ +\rangle$	0.79 ± 0.04		0.80 ± 0.03
$ -\rangle$	0.81 ± 0.04		0.77 ± 0.04
$ R\rangle$	0.75 ± 0.04		0.79 ± 0.04
$ L\rangle$	0.79 ± 0.04		0.82 ± 0.03

15. Zeilinger, A., Horne, M. A., Weinfurter, H. & Zukowski, M. Three-particle entanglements from two entangled pairs. *Phys. Rev. Lett.* **78**, 3031–3034 (1997).
16. Rarity, J. G. & Tapster, P. R. Three-particle entanglement from entangled photon pairs and a weak coherent state. *Phys. Rev. A* **59**, R35–R38 (1999).
17. Pan, J.-W., Simon, C., Brukner, C. & Zeilinger, A. Entanglement purification for quantum communication. *Nature* **410**, 1067–1070 (2001).
18. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
19. Bouwmeester, D. Bit-flip-error rejection in optical quantum communication. *Phys. Rev. A* **63**, R040301 (2001).
20. Hein, M., Eisert, J. & Briegel, H.J. Multi-party entanglement in graph states. *Phys. Rev. A* **69**, 06231 (2004).
21. Briegel, H. J. & Raussendorf, R. Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.* **86**, 910–913 (2001).
22. Kwiat, P. G. *et al.* New high intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
23. Kurtsiefer, C. *et al.* A step towards global key distribution. *Nature* **419**, 450 (2002).
24. De Martini, F., Buzek, V., Sciarrino, F. & Sias, C. Experimental realization of the quantum universal NOT gate. *Nature* **419**, 815–818 (2002).
25. Pan, J.-W., Gasparoni, S., Rupert, U., Weihs, G. & Zeilinger, A. Experimental entanglement purification of arbitrary unknown states. *Nature* **423**, 417–422 (2003).
26. Zukowski, M., Zeilinger, A. & Weinfurter, H. Entangling photons radiated by independent pulsed source. *Ann. NY Acad. Sci.* **755**, 91–102 (1995).
27. Pittman, T. B., Jacobs, B. C. & Franson, J. D. Probabilistic quantum encoder for single-photon qubits. *Phys. Rev. A* **69**, 042306 (2004).
28. Zukowski, M. & Kaszlikowski, D. Critical visibility for N-particle Greenberger-Horne-Zeilinger correlations to violate local realism. *Phys. Rev. A* **56**, R1682–R1685 (1997).
29. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
30. Pittman, T. B., Jacobs, B. C. & Franson, J. D. Demonstration of nondeterministic quantum logic operations using linear optical elements. *Phys. Rev. Lett.* **88**, 257902 (2002).

Acknowledgements This work was supported by the National Natural Science Foundation of China, the Chinese Academy of Sciences, the National Fundamental Research Program and the German Research Foundation (DFG).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.-W.P. (jian-wei.pan@physi.uni-heidelberg.de).

Cyclotron frequency shifts arising from polarization forces

James K. Thompson, Simon Rainville & David E. Pritchard

Research Laboratory of Electronics, MIT-Harvard Center for Ultracold Atoms, and Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

The cyclotron frequency of a charged particle in a uniform magnetic field B is related to its mass m and charge q by the relationship $\omega_c = qB/m$. This simple relationship forms the basis for sensitive mass comparisons using ion cyclotron resonance mass spectroscopy, with applications ranging from the identification of biomolecules¹ and the study of chemical reaction rates² to determinations of the fine structure constant of atomic spectra³. Here we report the observation of a deviation from the cyclotron frequency relationship for polarizable particles: in high-accuracy measurements of a single CO^+ ion, a dipole induced in the orbiting ion shifts the measured cyclotron frequency. We use this cyclotron frequency shift to measure non-destructively the quantum state of the CO^+ ion. The effect also provides a means to determine to a few per cent the body-frame dipole moment of CO^+ , thus establishing a method for measuring dipole moments of molecular ions for which few comparably accurate measurements exist^{4–6}. The general perturbation that we describe here affects the most precise mass comparisons attainable today^{7,8}, with applications including direct tests of Einstein's

mass–energy relationship⁹ and charge-parity-time reversal symmetry¹⁰, and possibly the weighing of chemical bonds⁷.

The cyclotron frequency shift reported here for an ion with electric polarizability α can be understood using the exaggerated microscopic picture shown in Fig. 1. As the molecule (with net charge indicated by +) moves on its cyclotron orbit (solid circle) with velocity $\mathbf{v}$ perpendicular to the magnetic field $\mathbf{B}$, the Lorentz force is experienced in its instantaneous rest frame as a motional electric field $\mathbf{E}_m = \mathbf{v} \times \mathbf{B}$. The motional electric field induces a dipole $\mathbf{d}_{\text{ind}} = \alpha \mathbf{E}_m$ (modelled here by the outer + and – charges), which points towards the centre of the cyclotron orbit. As the ion orbits, the orientation of the induced dipole adiabatically follows the motional electric field (as indicated by the two snapshots at times t and $t + \Delta t$). As a result, the two ends of the induced dipole move at slightly different speeds (as emphasized by the dashed lines). The differential velocity gives rise to a net additional Lorentz force on the particle that is also directed along the radial direction. The size of the induced dipole is proportional to the ion's velocity, which for the circular cyclotron motion is proportional to its cyclotron radius. A constant frequency shift then results because the size of the additional Lorentz force scales linearly with the cyclotron radius, just as the original Lorentz force does. Equivalently, the cyclotron frequency shift occurs because the induced dipole moves the centre of charge to a radius different from that of the centre of mass. The fractional frequency shift is then simply given by the ratio of the distance between the two centres and the cyclotron radius.

This polarization force also occurs for neutral particles, and can be generally modelled as a renormalization of the inertial mass of the particle in the plane perpendicular to the magnetic field. If a neutral particle accelerates, the induced dipole relaxes to a new equilibrium value, creating a transient Lorentz force that either opposes or enhances the change in velocity, acting like a changed inertial mass. The new effective mass is given simply by $m_\lambda = m(1 + \alpha_\lambda B^2/m)$ where m is the usual inertial mass, and α_λ is the electric polarizability of the particle. This expression for the effective mass can be derived by including a phenomenological polarization

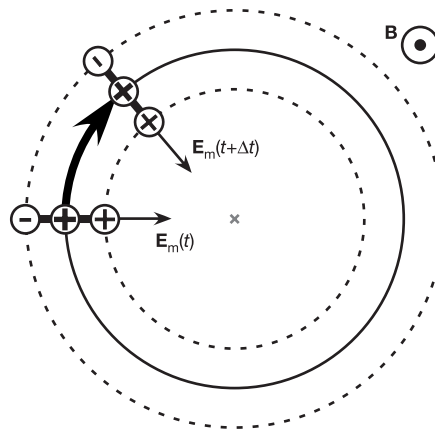


Figure 1 An exaggerated microscopic picture of the polarization force shift of the cyclotron frequency. As the molecule (with net charge indicated by +) moves on its cyclotron orbit (solid circle) perpendicular to the magnetic field $\mathbf{B}$, the magnetic field is experienced in its instantaneous rest frame as a motional electric field $\mathbf{E}_m$. The motional electric field induces a dipole (modelled here by the outer + and – charges), which points towards the centre of the cyclotron orbit. As the ion orbits, the orientation of the induced dipole adiabatically follows the motional electric field (as indicated by the two snapshots at times t and $t + \Delta t$). As a result, the two ends of the induced dipole move at slightly different speeds (as emphasized by the dashed lines) thus generating an additional Lorentz force. The orbital frequency then has to adjust in order to compensate for the additional force.

energy in the lagrangian, as was first pointed out in the context of neutral atom interferometry¹¹. Since the polarizability may depend on the quantum state of the particle, it carries a label λ to specify the state. A key result here is that the effective mass m_λ , as determined from the measured cyclotron frequency $\omega_c^\lambda = qB/m_\lambda$, depends on the quantum state of the particle through the polarizability. For the present work, the fractional modification of the cyclotron frequency with respect to that for a particle of the same mass and charge but zero polarizability is $\Delta\omega_c^\lambda/\omega_c = -\alpha_\lambda B^2/m \approx 10^{-9}$. Note that this is a fundamental shift of the cyclotron frequency, in contrast to shifts arising from experimental artefacts (such as the image charge shift or the quadrupole electric field used to confine the ions along the magnetic field lines during the measurement). This cyclotron frequency shift has not been previously recognized or observed, but it is crucial to consider for a precision better than 10^{-9} , which is now being approached by several groups^{7,8,10,12}.

Experimentally, magnetic field noise limits the ability to observe the expected changes in the cyclotron frequency $\Delta\omega_c^\lambda/\omega_c$. Instead, we measure the ratio of cyclotron frequencies for two different ions (only one of which is polar). For the best discrimination against magnetic and other sources of noise on the ratio, we recently developed a technique to confine the two different ions in a Penning trap and simultaneously compare their cyclotron frequencies, as is described in ref. 7. With this novel technique, we can obtain mass ratios with fractional accuracies below 10^{-11} .

Quantum-state-dependent polarization shifts of the cyclotron frequency are clearly observable in the time evolution of the cyclotron frequency ratio of a single CO^+ and a single N_2^+ (see Fig. 2a). The ratio changes by $\sim 10^{-9}$ in discrete jumps between approximately three values, remaining stable for many hours

between the jumps. That the variation was of the CO^+ cyclotron frequency, and not of the N_2^+ , was confirmed by directly observing the individual cyclotron frequencies of both the CO^+ and N_2^+ using a magnetometer to remove external magnetic field fluctuations—a technique sufficient for coarsely identifying jumps of order 10^{-9} . Additionally, the measured cyclotron frequency ratio of the symmetric molecules N_2^+ and $^{13}\text{C}_2\text{H}_2^+$ (which have much smaller polarizabilities than the asymmetric CO^+) did not manifest any jumps^{7,13}.

We considered and eliminated many possible experimental artefacts as being the source of the jumps. We developed techniques to measure precisely the orbits of the simultaneously confined ions and found no correlation between the jumps and changes in their orbits (which coupled with our magnetic field inhomogeneity might give rise to jumps of the observed magnitude). Also, the size and rate of the jumps does not vary with ion-ion separation, counter-indicating Coulomb interactions as a source of the jumps. Combined with numerous other checks, we feel confident that the observed jumps in the cyclotron frequency ratio are not experimental artefacts of our technique for accomplishing simultaneous cyclotron frequency comparisons.

The ratio jumps because the cyclotron frequency of the CO^+ molecule changes when it absorbs or emits black-body radiation and moves from one quantum rotational state to another. To interpret the observed data, one must first know that the CO^+ and N_2^+ molecules are trapped for many days, giving the CO^+ time to come to thermal equilibrium with the 4.2 K environment of the trap. At this temperature, the CO^+ is in the ground electronic and vibrational state and spends the majority of its time in the lowest few rotational levels. The states with total angular momentum $N = 0$ and 1 will be shown to give rise to the three distinct cyclotron frequencies observed.

To determine the best basis in which to calculate the expected cyclotron frequency shifts, we first consider the hierarchy of energy scales. The CO^+ ion possesses an unpaired electron whose spin strongly couples to the 8.5 T magnetic field resulting in a set of independent spin up and spin down manifolds. A much smaller spin-rotation coupling $\hbar\gamma_e\mathbf{S}\cdot\mathbf{N}$, with coupling constant $\gamma_e/2\pi = 0.57$ GHz, fixes the quantization axis for the rotational states along the spin quantization axis, that is, the magnetic field direction ($\hbar$ is the reduced Planck's constant)¹⁴. The energy splitting between rotational levels N and $N + 1$ is given by $E_{N,N+1} = 4\pi\hbar B_0(N + 1)$, with the rotational constant $B_0 = 58.983$ GHz (ref. 14).

In the ion's instantaneous rest frame, the motional electric field arising from the cyclotron motion appears to rotate in the x - y plane at the cyclotron frequency $\omega_c/2\pi \approx 4.67$ MHz. The interaction is quasi-static because $E_{N,N+1}/\hbar$ and γ_e are both much larger than ω_c ; hence, the induced dipole adiabatically rotates with the motional electric field. In addition, a typical cyclotron radius of $\rho_c = 75$ μm gives a motional electric field $E_m = 190$ V cm^{-1} , and polarization

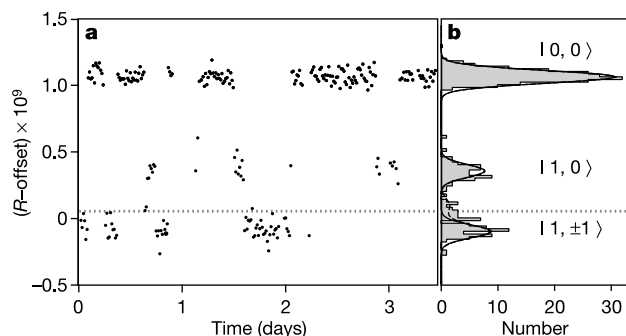


Figure 2 The cyclotron frequency ratio $R = \omega_c[\text{N}_2^+]/\omega_c[\text{CO}^+]$. **a**, Measured versus time, and **b**, a histogram of all measured values. The subtracted offset is our previously measured ratio $R = 0.999\,598\,887\,572$ (determined by alternately trapping the two ions²⁰). The ratio changes when the CO^+ ion absorbs or emits a black-body photon and jumps to a quantum rotational state with a different electric polarizability. The many hour lifetimes of the states and the fraction of time spent in each state are in reasonable agreement with the calculated values (see Table 1). The spectrum of measured ratios has three main components, as is made more clear by the histogram in **b**, which includes additional data not shown in **a**. The separation between the dominant three peaks allows the rotational state $|N, M_M\rangle$ of the CO^+ molecule to be non-destructively determined from a single ratio measurement. The curves are fits to determine the molecular dipole moment μ and the cyclotron frequency ratio unperturbed by polarization forces. The solid line includes only contributions from $N = 0$ and 1 (reduced chi-square $\chi_r^2 = 1.1$), while the dashed line also includes small contributions from the $N = 2$ rotational states ($\chi_r^2 = 0.8$). The fits have the following adjustable parameters: the unperturbed cyclotron frequency ratio R , the dipole moment μ , a common gaussian width arising from the measurement noise, and a separate amplitude for each gaussian. The grey dashed line indicates the final unperturbed ratio R as reported in the text. The quoted errors on the parameters μ and R were conservatively increased to reflect small variations resulting from, for example, changing the binning and including only two of the three states at a time in the fit.

Table 1 Calculated and measured properties for CO^+

N	M_N	α_λ (10^{-38} C m ² V ⁻¹)	$\Delta\omega_c^\lambda/\omega_c$ (10^{-12})		τ (h)		Occupation	
			Th.	Exp.	Th.	Exp.	Th.	Exp.
0	0	63.14	-988	-1008(30)	6.3	7(3)	0.54(6)	0.65
1	0	18.96	-297	-302(9)	2.4	-	0.14(9)	0.16
1	± 1	-9.462	148	151(5)	2.4	4(2)	0.28(11)	0.19
2		<4.5	<70	-	0.04	-	0.04	-

The theoretical (Th.) values are calculated using $\mu = 1.015e a_0$. The experimental (Exp.) fractional cyclotron frequency shifts $\Delta\omega_c^\lambda/\omega_c$ are from the fit shown in Fig. 2b. The calculated lifetimes τ include both spontaneous and thermally induced transition rates at 4.2 K. The occupation is the fraction of the total time spent in each state as calculated from the normalized Boltzmann weights at 4.2 K. The uncertainties on the theoretical occupation values (given in parentheses) account for the finite observation time of approximately 4 days compared to the theoretical transition rates. Values indicated by - are not quoted owing to insufficient data.

energies of only $V_p = -\frac{1}{2}\alpha_\lambda E_m^2 \leq 2\pi \times 0.17 \text{ MHz}$. Hence, we are justified in using the d.c. polarizability calculated using standard second-order time-independent perturbation theory, in which the molecular dipole moment squared enters as an overall scale factor¹⁵.

For the pertinent two lowest rotational levels, the above model predicts three distinct values of the polarizability. These can be understood qualitatively as follows (see Table 1 for the quantitative results.) The motional electric field, lying in the x - y plane, only couples states $|N, M_N\rangle$ and $|N', M'_N\rangle$ for which $N' - N = \pm 1$ and $M'_N - M_N = \pm 1$. Thus, the state $|0, 0\rangle$ and pair of states $|1, \pm 1\rangle$ repel one another leading to opposite signs of the polarizability between the ground state and these two excited states. Meanwhile, the state $|1, 0\rangle$ does not interact with the ground state, but is repelled downwards by the pair of states $|2, \pm 1\rangle$ so that its polarizability has the same sign as $|0, 0\rangle$, but a smaller magnitude owing to the larger energy denominator and smaller Clebsch-Gordon coefficients.

The calculated free space lifetimes of the states (including thermally induced transition rates) are in good agreement with the observed lifetimes (Table 1). In Fig. 2a, the observed transitions between $|1, 0\rangle$ and $|1, \pm 1\rangle$ (which would violate parity) are explained by pairs of transitions to and back from a state $|2, M_N\rangle$. These transitions are not clearly resolved owing to the short lifetime of 3 min for $N = 2$ states, whereas the ratio is averaged over the typical 5 min comparison interval.

The above model quantitatively accounts for the observed distribution of cyclotron frequency ratios, and can be used to determine accurately the molecular dipole moment of CO^+ and the unperturbed mass ratio $R = M[\text{CO}^+]/M[\text{N}_2^+]$. A histogram is made of the number of times each cyclotron frequency ratio is measured (Fig. 2b.) A simultaneous fit to the dominant three peaks, in which the molecular dipole moment acts as a scale factor while the ratio of the intervals between the peaks is held fixed by theory, yields a good reduced chi-squared of $\chi^2_\nu = 1.1$ (see solid curve of Fig. 2b and legend for details). The fitted molecular dipole moment is $\mu = 1.025(15)e a_0$, in agreement with the theoretical prediction¹⁶ $\mu = 1.015e a_0$, where e is the elementary charge, a_0 is the Bohr radius, and the number in parentheses indicates the standard uncertainty (or standard deviation of the mean) on the last two digits.

The dipole moments of molecular ions are difficult to measure, and only a few have been determined from the rotational Zeeman effect by laser spectroscopy⁴⁻⁶. The technique demonstrated here offers an alternative route to important physical chemistry data, as well as being both conceptually simple and providing an accuracy comparable to the best spectroscopic determinations. Also, because only a single ion is necessary, this technique is applicable to highly reactive molecules or those for which large enough quantities cannot be synthesized for laser spectroscopy.

Because of better statistics and discrimination, we can measure the cyclotron frequency ratio $R_{0,0}$ of the ground state with an uncertainty of 7×10^{-12} , which includes the statistical and systematic errors from ion-ion interaction and trap field imperfections⁷. Unfortunately, the accuracy of the mass ratio we extract, $R = M[\text{CO}^+]/M[\text{N}_2^+] = 0.999\,598\,887\,627(15)$, is limited by our knowledge of μ . Still, the fractional uncertainty of 1.5×10^{-11} makes this one of the three most precisely known mass ratios. A measurement or calculation of the electric dipole moment with fractional accuracy of 0.5% would be needed to determine the unperturbed mass ratio from the ground-state cyclotron frequency ratio with the same accuracy as $R_{0,0}$ from $R = M[\text{CO}^+]/M[\text{N}_2^+] = 0.999\,598\,888\,635\,(7) - (959 \times \mu^2 \times 10^{-12})$ where μ is expressed in $e a_0$. After correcting for the mass of the missing electrons¹⁷, and ionization and molecular binding energies¹⁸, we obtain a neutral mass difference $2 \times M[^{14}\text{N}] - M[^{16}\text{O}] - 12\text{ u} = 0.011\,233\,389\,32(42)\text{ u}$ in unified atomic mass units. This mass comparison is metrologically important because it determines the neutral atomic masses $M[^{14}\text{N}]$ and $M[^{13}\text{C}]$ to fractional accuracies of $\sim 2 \times 10^{-11}$

when combined with previously reported mass comparisons^{7,8,19}.

The above mass ratio differs from our previously measured value, obtained by alternately trapping and measuring the cyclotron frequencies of the two ions²⁰, by less than the previous uncertainty of 7×10^{-11} . We believe that our group's previous measurements^{3,20} do not significantly suffer from any systematic error from polarization shifts because the rotational state, and hence the systematic error, was randomized each time a new ion was created by electron impact ionization. Moreover, the ions did not have time to relax to the ground rotational state (the only state possessing a large polarizability) during the subsequent ~ 10 min over which the cyclotron frequency was measured.

The cyclotron frequency shift reported here is a very general phenomenon affecting all cyclotron frequency comparisons involving polarizable particles. For instance, the most precise test of charge-parity-time (CPT) reversal symmetry using baryons¹⁰ was performed by comparing the charge to mass ratios of the proton p and antiproton $\bar{p}$. The measurement was actually performed by comparing the cyclotron frequencies of H^- and $\bar{p}$. Correcting for the polarization force shift of the H^- ion's cyclotron frequency^{10,21} using $\alpha[\text{H}^-] = 0.34 \times 10^{-38} \text{ C m}^2 \text{ V}^{-1}$ and $B = 5.9 \text{ T}$, the reported ratio changes from $(q[\bar{p}]/m[\bar{p}])/(q[p]/m[p]) + 1 = 0.9(9) \times 10^{-10}$ to a corrected value of $1.6(9) \times 10^{-10}$. Further, uncertainties in the size of the polarization shift of the SiH^+ and SH^+ molecules are expected to limit the accuracy of the mass comparisons for a new direct test of Einstein's mass-energy relationship $E = mc^2$, achieved by comparing the energy of the γ -rays emitted in a neutron capture process to the measured mass shift between the isotopes (ref. 9, and J.K.T., S.R., E.G. Myers and D.E.P., manuscript in preparation).

The cyclotron frequency shift presented here can be much larger in other systems. For scale, we consider a large linear molecule with $\mu = 5e a_0$, $M = 280 \text{ u}$, and the same linear mass density as CO^+ , measured in the currently attainable magnetic field of $B = 20 \text{ T}$. The fractional cyclotron frequency shift of the ground rotational state would be $\Delta\omega/\omega_c \approx 10^{-5}$, gigantic by the standards of state-of-the-art cyclotron frequency comparisons, and within reach of the higher-performance ion cyclotron resonance experiments used for physical or biological chemistry¹. However, for such a hypothetical molecule, the induced dipole approaches the molecular dipole moment at a cyclotron radius of only $\rho_c^{\text{critical}} \approx 10 \mu\text{m}$. Above this critical value, the cyclotron frequency shift should roughly fall off from its small orbit value as $\rho_c^{\text{critical}}/\rho_c$. Additionally, such a large molecule in thermal equilibrium would occupy much higher rotational states ($N \approx 25$ being the most likely occupied rotational level at 4 K) possessing smaller polarization shifts than that of the ground state. Therefore, this polarization force might manifest itself mostly as a broadening of the cyclotron resonance in such experiments.

This newly discovered cyclotron frequency shift will need to be carefully accounted for in future high-precision cyclotron frequency comparisons. Also, the demonstrated quantum non-demolition measurement of the rotational state of a single polar molecular could prove useful for high-precision single-molecule spectroscopy—a particularly interesting idea for possible electron dipole moment searches using polar molecules²². □

Received 4 April; accepted 25 May 2004; doi:10.1038/nature02682.

1. Marshall, A. G., Hendrickson, C. L. & Shi, S. D. Scaling MS plateaus with FTICR MS. *Anal. Chem.* **74**, 252A–259A (2002).
2. Heninger, M. et al. Successive reactions of iron carbonyl cations with methanol. *Int. J. Mass Spectrom. Ion Process.* **199**, 267–285 (2000).
3. Bradley, M. P., Porto, J. V., Rainville, S., Thompson, J. K. & Pritchard, D. E. Penning trap measurements of the masses of ^{133}Cs , ^{87}Rb , ^{85}Rb , and ^{23}Na with uncertainties $\leq 0.2 \text{ ppb}$. *Phys. Rev. Lett.* **83**, 4510–4513 (1999).
4. Havenith, M., Zwart, E., Meerts, W. L. & ter Meulen, J. J. Determination of the electric dipole moment of HN_2^+ . *J. Chem. Phys.* **93**, 8446–8451 (1990).
5. Laughlin, K. B., Blake, G. A., Cohen, R. C., Hovde, D. C. & Saykally, R. J. Determination of the dipole moment of ArH^+ from the rotational Zeeman effect by tunable far-infrared laser spectroscopy. *Phys. Rev. Lett.* **58**, 996–999 (1987).

6. Linnartz, H., Havenith, M., Zwart, E., Meerts, W. L. & ter Meulen, J. J. Determination of the electric-dipole moment of KrH^+ . *J. Mol. Spectrosc.* **153**, 710–717 (1992).
7. Rainville, S., Thompson, J. K. & Pritchard, D. E. An ion balance for ultra-high-precision atomic mass measurements. *Science* **303**, 334–338 (2004).
8. VanDyck, R. S., Zafonte, S. L. & Schwinberg, P. B. Ultra-precise mass measurements using the UW-PTMS. *Hyperfine Interact.* **132**, 163–175 (2001).
9. Greene, G. L., Dewey, M. S., Kessler, E. G. & Fischbach, E. Test of special relativity by a determination of the Lorentz limiting velocity — does $E = mc^2$? *Phys. Rev. D* **44**, R2216–R2219 (1991).
10. Gabrielse, G. *et al.* Precision mass spectroscopy of the antiproton and proton using simultaneously trapped particles. *Phys. Rev. Lett.* **82**, 3198–3201 (1999).
11. Wei, H. Q., Han, R. S. & Wei, X. Q. Quantum phase of induced dipoles moving in a magnetic field. *Phys. Rev. Lett.* **75**, 2071–2073 (1995).
12. Bergstrom, I. *et al.* SMILETRAP — A Penning trap facility for precision mass measurements using highly charged ions. *Nucl. Instrum. Methods Phys. Res. A* **487**, 618–651 (2002).
13. Jagod, M. F. *et al.* Infrared spectroscopy of carbo-ions. VI. C–H stretching vibration of the acetylene ion C_2H_2^+ and isotopic species. *J. Chem. Phys.* **97**, 7111–7123 (1992).
14. Dixon, T. A. & Woods, R. C. Microwave absorption spectrum of the CO^+ ion. *Phys. Rev. Lett.* **34**, 61–63 (1975).
15. Townes, C. H. & Schawlow, A. L. *Microwave Spectroscopy* (Dover Pub. Inc., New York, 1975).
16. Martin, P. A. & Feher, M. CASSCF calculations of the multipole moments and dipole polarizability functions of the $X^2\Sigma^+$ and $A^2\Pi$ states of CO^+ . *Chem. Phys. Lett.* **232**, 491–496 (1995).
17. Beier, T. *et al.* New determination of the electron's mass. *Phys. Rev. Lett.* **88**, 011603 (2002).
18. Huber, K. P. & Herzberg, G. (data prepared by Gallager, J. W. & Johnson, R. D.) “*Constants of Diatomic Molecules*”, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69* (eds Lindstrom, P. J. & Mallard, W. G.) (National Institute of Standards and Technology, Gaithersburg, MD, 2003); (<http://webbook.nist.gov>) (2003).
19. Audi, G., Wapstra, A. H. & Thibault, C. The AME2003 atomic mass evaluation (II). Tables, graphs and references. *Nucl. Phys. A* **729**, 337–676 (2003).
20. DiFilippo, F., Natarajan, V., Boyce, K. R. & Pritchard, D. E. Accurate atomic masses for fundamental metrology. *Phys. Rev. Lett.* **73**, 1481–1484 (1994).
21. Bhatia, A. K. & Drachman, R. J. Polarizability of helium and the negative hydrogen ion. *J. Phys. B* **27**, 1299–1305 (1994).
22. Kozlov, M. G. & DeMille, D. Enhancement of the electric dipole moment of the electron in PbO . *Phys. Rev. Lett.* **89**, 133001 (2002).

Acknowledgements We thank E.G. Myers, G.J. Sussman, J. Wisdom and W. Ketterle for useful discussions. This work was supported by the National Science Foundation and a National Institutes of Standards and Technology Precision Measurement Grant. S.R. acknowledges support from the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.K.T. (jkthomps@alum.mit.edu).

Single-crystal metallic nanowires and metal/semiconductor nanowire heterostructures

Yue Wu^{1*}, Jie Xiang^{1*}, Chen Yang¹, Wei Lu¹ & Charles M. Lieber^{1,2}

¹Department of Chemistry and Chemical Biology; and

²Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

* These authors contributed equally to this work

Substantial effort has been placed on developing semiconducting carbon nanotubes^{1–3} and nanowires⁴ as building blocks for electronic devices—such as field-effect transistors—that could replace conventional silicon transistors in hybrid electronics or lead to stand-alone nanosystems^{4,5}. Attaching electric contacts to individual devices is a first step towards integration, and this step has been addressed using lithographically defined metal electrodes^{1–4,6–8}. Yet, these metal contacts define a size scale that is much larger than the nanometre-scale building blocks, thus limiting many potential advantages. Here we report an integrated contact and interconnection solution that overcomes this size constraint through selective transformation of silicon nanowires into metallic nickel silicide (NiSi) nanowires. Electrical measurements

show that the single crystal nickel silicide nanowires have ideal resistivities of about $10\ \mu\Omega\text{ cm}$ and remarkably high failure-current densities, $>10^8\text{ A cm}^{-2}$. In addition, we demonstrate the fabrication of nickel silicide/silicon (NiSi/Si) nanowire heterostructures with atomically sharp metal–semiconductor interfaces. We produce field-effect transistors based on those heterostructures in which the source–drain contacts are defined by the metallic NiSi nanowire regions. Our approach is fully compatible with conventional planar silicon electronics and extendable to the 10-nm scale using a crossed-nanowire architecture.

Our focus on NiSi nanowires is motivated by previous investigations of metal silicides, which can exhibit low resistivity, compatibility with conventional silicon manufacturing, and the ability to form ohmic contacts to both p- and n-type silicon⁹. We prepared NiSi nanowires using an approach (Fig. 1a) involving deposition of nickel metal onto single-crystal Si nanowires^{10,11}, solid-state reaction at 550 °C to form NiSi, followed by removal of remaining metal by wet etching (see Methods). Low-resolution transmission electron microscopy (TEM) studies (Fig. 1b) of materials prepared in this

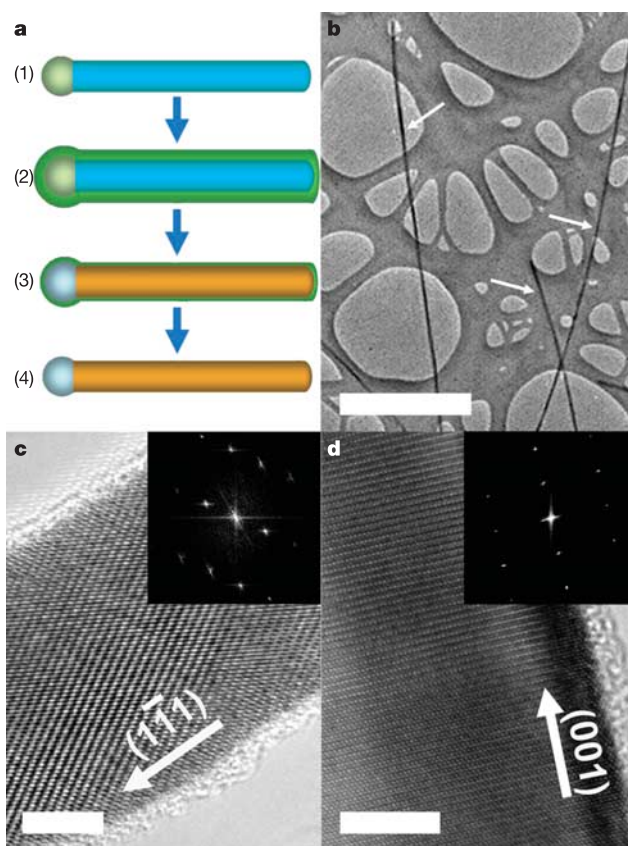


Figure 1 Preparation and structural characterization of single-crystal NiSi nanowires. **a**, Preparation of single-crystal NiSi nanowires. (1) Si nanowires (blue) of uniform diameter are (2) coated with Ni metal (green) to a total thickness comparable to the Si nanowire diameter, (3) reacted at 550 °C to form NiSi nanowires (brown), and then (4) etched to remove any excess Ni metal. **b**, TEM image of three NiSi nanowires prepared using Si nanowires with an average diameter of about 20 nm. The NiSi nanowires are highlighted by white arrows; scale bar is 1 μm . **c** and **d**, High-resolution TEM images of single-crystal NiSi nanowires. **c**, TEM image of a 20-nm NiSi nanowire prepared using Si nanowires of average diameter 20 nm. The white arrow indicates the growth front of the nanowire. Inset, two-dimensional Fourier transform of the image showing the $[10\bar{1}]$ zone axis of NiSi. **d**, Image of a 32-nm NiSi nanowire prepared using Si nanowires of average diameter 30 nm. The white arrow indicates the growth front of the nanowire. Inset, two-dimensional Fourier transform of the image depicting the $[2\bar{1}0]$ zone axis of NiSi. The scale bars in **c** and **d** are 5 nm.

way show nanowires that have uniform diameters and contrast, which are properties indicative of single-crystal structures, and lengths of tens of micrometres. Analysis of samples prepared from Si nanowires with diameters of 20.3 ± 2.3 nm yielded materials with diameters of 22.8 ± 3.4 nm. The increase in average diameter agrees with the value of 22.4 nm, expected for conversion of Si to NiSi on the basis of different unit-cell volumes¹². In addition, energy dispersive X-ray spectroscopy (EDS) measurements show that the Ni:Si atom ratio in the nanowires is 1.03:1, as expected for NiSi.

High-resolution TEM studies (Figs 1c and d) demonstrate clearly that the nanowires are single-crystal NiSi. The reciprocal lattice peaks, which were obtained from two-dimensional Fourier transforms (2DFT) of the lattice-resolved images (insets to Figs 1c and d) can be indexed to the orthorhombic structure of NiSi¹² with the zone axes along the [101] and [210] directions, respectively. This analysis enables assignment of the growth front of NiSi to be the (111) and (001) planes for the 20- and 30-nm-diameter nanowires in Figs 1c and d, respectively. The different NiSi nanowire growth directions may be explained by different growth directions of the starting silicon nanowires¹¹, although future work will be required to quantify this relationship. These high-resolution data also show that the NiSi nanowires have smooth surfaces with very little (<1 nm) amorphous coating. Taken together, these results demonstrate that our approach yields uniform conversion of single-crystal Si nanowires to single-crystal NiSi nanowires. Lastly, similar studies with germanium (Ge) nanowires show that metallic single-crystal NiGe nanowires can also be prepared by our approach (Y.W., J.X. & C.M.L., unpublished results).

Representative electrical transport data recorded on a 29-nm-diameter NiSi nanowire (Fig. 2a) show linear current (I) versus voltage (V) behaviour with two- and four-terminal resistances of 886 and 184 Ω , respectively. The current levels observed from 30-nm NiSi nanowires typically exceed 100 μ A at 100 mV bias, are quite remarkable in comparison to the starting Si nanowires, and are clearly indicative of the expected metallic nature of NiSi. Indeed, calculation of the resistivity on the basis of the four-terminal data yields a value of $9.5 \mu\Omega$ cm, close to the value of $10 \mu\Omega$ cm for NiSi single crystals¹³. These low-resistivity values have been observed for NiSi nanowires with diameters from about 15 to 45 nm, and thus demonstrate that these single-crystal wires can be scaled to ultrasmall dimensions without degradation of properties. Temperature-dependent measurements (inset to Fig. 2a) further show that the nanowire resistance decreases monotonically down to about 30 K and then saturates as expected for a metal. The similarity of NiSi nanowire resistivities to the bulk value, the temperature dependence of the resistance, and the scaling of resistance with length (to 6μ m) are indicative of diffusive transport. Using a previously reported carrier density value for NiSi¹³ we estimate the scattering mean free path to be of the order of 5 nm. Hence, it should be possible to retain the attractive metallic transport in NiSi nanowires down to the sub-10-nm-diameter scale, and also to study the effects of fundamental properties, such as dephasing, in a pure metal regime that should be accessible from recently reported molecular-scale silicon nanowires¹¹ using our synthetic approach.

We have also characterized the maximum transport current for the metallic NiSi nanowires. For example, the 29-nm-diameter nanowire discussed above could carry a current of 1.84 mA before failure (Fig. 2b). Notably, this yields a maximum current density, $J_{\max}$, of 3×10^8 A cm⁻². This large value of $J_{\max}$ is reproducible, and is approximately independent of nanowire diameter. For example, a nanowire of diameter 40–45 nm exhibited a failure current of 5.5 mA, and $J_{\max} > 3 \times 10^8$ A cm⁻². In general, failure occurs in the middle of the nanowires, which is where the peak temperature is expected to occur, owing to dissipative self-heating; this suggests the breakdown mechanism is due to melting. The high $J_{\max}$ values can be attributed to the single-crystal structures, which preclude energy dissipation and void diffusion¹⁴ at the grain boundary and defect

sites that are common in lithographically defined wires.

The $J_{\max}$ for the NiSi nanowires is comparable to the best values— 10^9 A cm⁻²—reported for single-walled carbon nanotubes¹⁵. We believe this finding is important because the high $J_{\max}$ values are achieved in essentially every NiSi nanowire, without the need to find and connect the metallic versus semiconducting carbon nanotubes. In addition, the maximum current can be scaled through the growth of specific-diameter NiSi nanowires. The $J_{\max}$ for the NiSi nanowires is also about two orders of magnitude larger than that observed in lithographically defined noble metal lines¹⁶. This suggests an immediate advantage of NiSi nanowires over lithographically defined nanoscale metal lines: for example, they could be used as local- and intermediate-level interconnects to nanowire field-effect transistors that might be key components in hybrid nanoelectronics.

Importantly, the approach outlined above can also be used to transform selectively single-crystal silicon nanowires to produce NiSi/Si heterostructures and superlattices. To demonstrate this

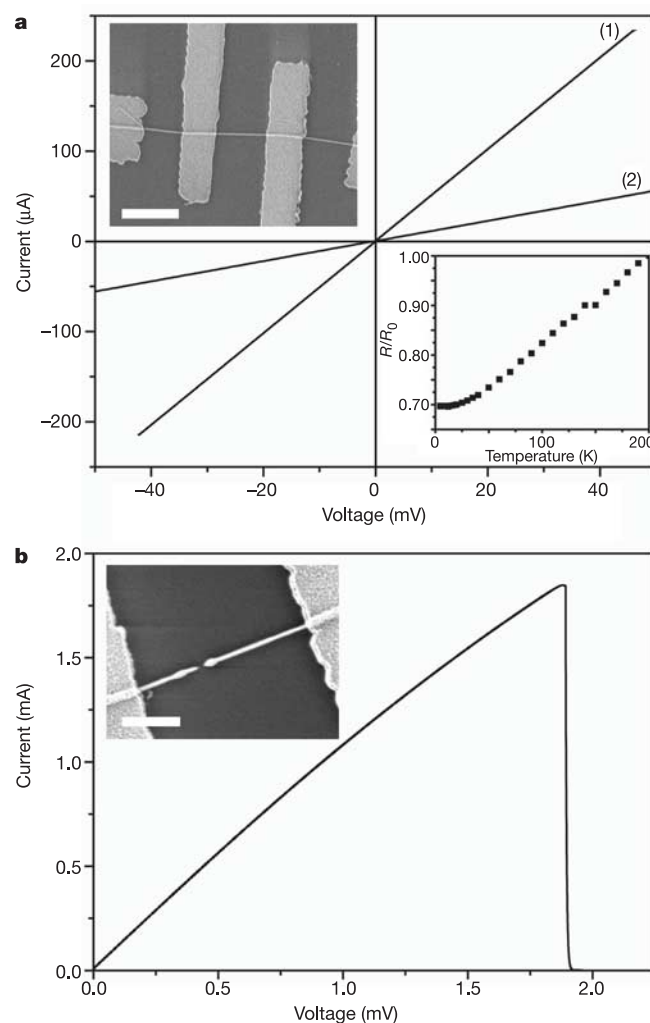


Figure 2 Transport measurements on individual single-crystal NiSi nanowires. **a**, Current versus voltage curves recorded on a 29-nm NiSi nanowire, with line (1) and line (2) corresponding to four- and two-terminal measurements, respectively, taken from the device shown in the SEM image (inset, upper left). The scale bar in the image is 1 μ m. Lower inset, temperature-dependent normalized resistance obtained from a four-terminal device; the resistance, R , is normalized by the value at 200 K, R_0 . **b**, I - V data recorded at large applied voltages. The rapid drop at about 1.8 V corresponds to the failure point of this nanowire. Inset, SEM image of the nanowire after breakdown. The image highlights the break close to the middle of the nanowires. Scale bar, 500 nm.

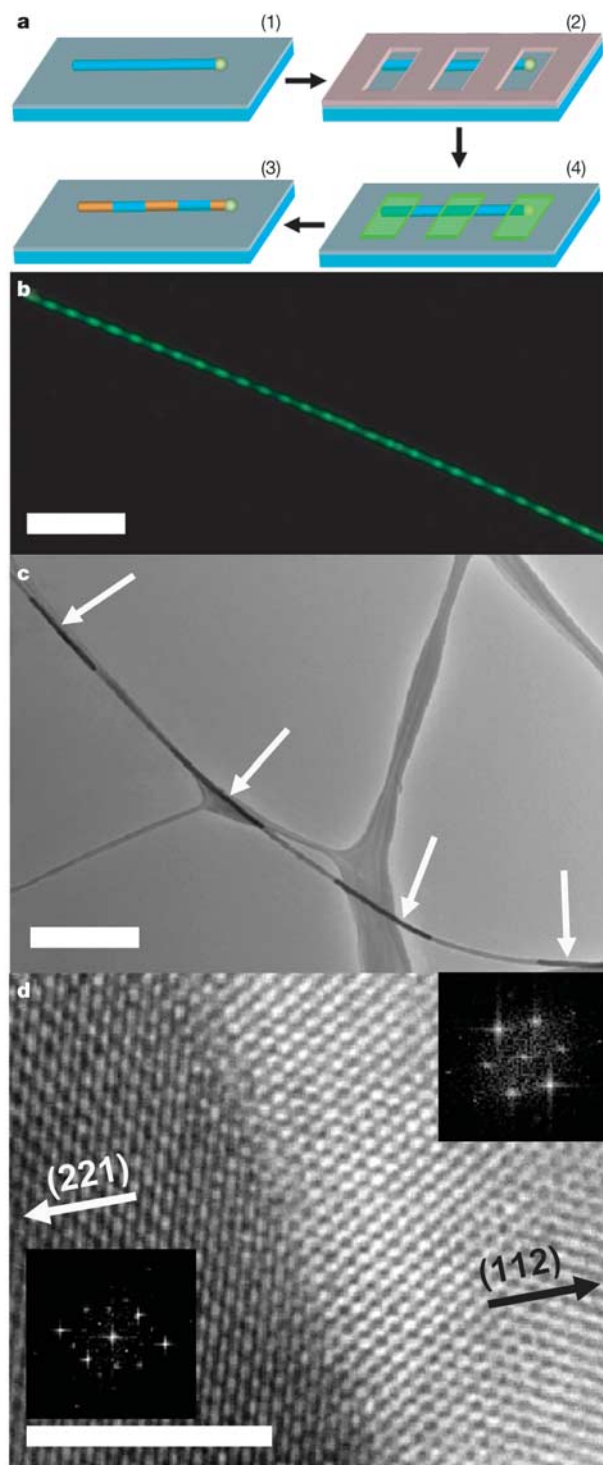


Figure 3 Fabrication and structural characterizations of NiSi/Si nanowire heterostructures and superlattices. **a**, Fabrication of NiSi/Si nanowire heterostructures and superlattices. (1) Si nanowires (blue) dispersed on a substrate are (2) coated with photoresist (grey) and lithographically patterned, (3) selectively coated with Ni metal (green) to a total thickness comparable to the Si nanowire diameter, and (4) reacted at 550 °C to form NiSi nanowires. **b**, Dark-field optical image of a single NiSi/Si nanowire heterostructure. The bright green segments correspond to silicon and the dark segments to NiSi. Scale bar is 10 μm . **c**, TEM image of a NiSi/Si heterostructured nanowire. The bright segments of the nanowire correspond to silicon and the dark segments, which are highlighted with arrows, correspond to NiSi. Scale bar is 1 μm . **d**, High-resolution TEM image of the junction between NiSi and Si showing an atomically abrupt interface. Insets, two-dimensional Fourier transforms of the image depicting the [110] and [111] zone axes of NiSi and Si, respectively, where the arrows highlight the growth fronts of the NiSi (221) and Si (112). Scale bar is 5 nm.

concept (Fig. 3a) we patterned nickel metal onto silicon nanowires and then transform the nickel-containing regions to NiSi (see Methods). A dark-field optical image of a nanowire patterned in this way using 1- μm -wide nickel regions on a 2- μm pitch (Fig. 3b) exhibits periodic variations in contrast extending over the full length of the 65- μm -long nanowire. Analysis of the image shows

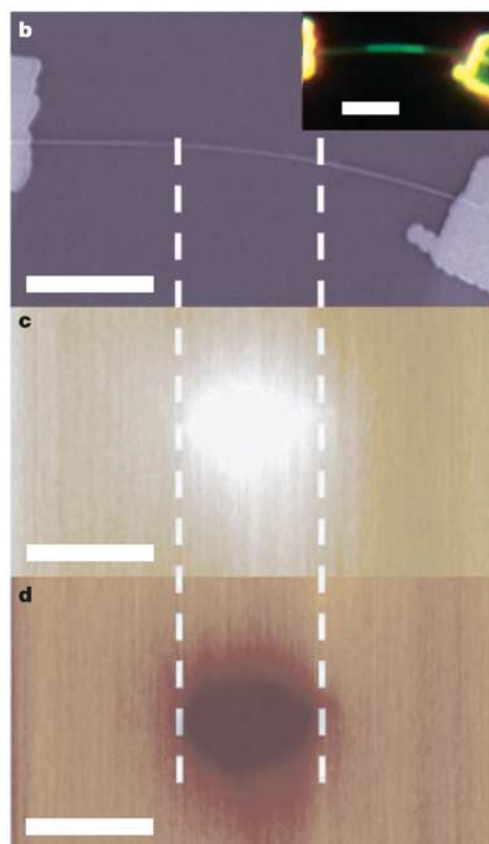
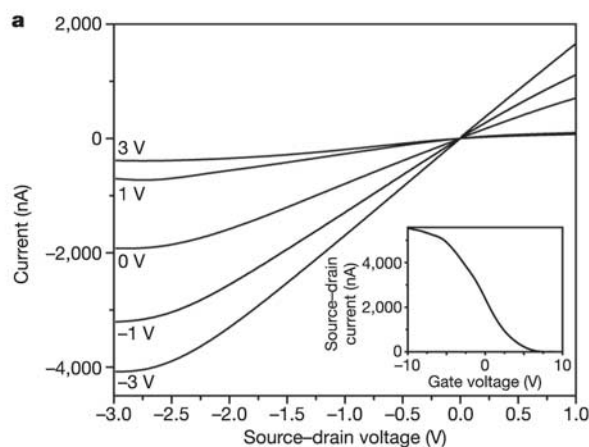


Figure 4 Transport properties of a NiSi/p-Si/NiSi heterojunction field-effect transistor. **a**, I versus V_{sd} , exhibiting the typical characteristics of a NiSi/p-Si/NiSi heterojunction transistor. Inset, gate sweep obtained from the same device in the saturation regime with $V_{\text{sd}} = -3 \text{ V}$. The device was fabricated using a 30-nm diameter p-Si nanowire (dopant density $\sim 1 \times 10^{18} \text{ cm}^{-3}$) on a heavily doped silicon substrate with 600 nm of thermal oxide; the channel length is 3 μm . **b**, SEM image of a field-effect transistor device fabricated using a NiSi/p-Si/NiSi nanowire heterojunction. Inset, dark-field optical image of the same device, where the bright green segment corresponds to silicon and the dark segments to NiSi. Scale bars, 3 μm . **c** and **d**, SGM images show reduced conductance with +9 V (Fig. 4c) and enhanced conductance with -9 V gate voltage (Fig. 4d) on the atomic force microscope tip. The dotted lines mark the interface between the NiSi and p-Si regions. Scale bars, 3 μm .

that the average lengths of the Si and NiSi regions are both 1 μm , and in good agreement with the width and pitch of nickel metal deposited on the nanowire during fabrication.

TEM images of similar NiSi/Si nanowire heterostructures (Fig. 3c) show a similar periodic variation in contrast that is consistent with NiSi (dark) and Si (light) materials within the heterostructure. This assignment was confirmed by EDS analysis: the dark and light regions corresponded to an Ni:Si ratio of 1.02:1.00 and pure Si, respectively. Analysis of the images also shows that the average lengths of the Si and NiSi regions are 0.93 ± 0.10 and 1.03 ± 0.11 μm , respectively. The good agreement with the expected pattern demonstrates clearly that our approach enables spatially controlled transformation of silicon to metallic NiSi nanowire heterostructures. Notably, detailed examination of NiSi/Si heterostructure by high-resolution TEM (Fig. 3d) shows that the transformation yields an atomically abrupt interface (irrespective of any axial diffusion). The reciprocal lattice peaks, which were obtained from two-dimensional Fourier transforms of the lattice-resolved image (insets to Fig. 3d) correspond to the [110] and [111] zone axes of NiSi and Si, with (221) and (112) nanowire growth fronts, respectively.

The atomically sharp metal–semiconductor interfaces produced in these nanowire heterostructures have the potential to yield a range of precisely defined electronic devices and device arrays on individual nanowires. To explore this opportunity we have prepared field-effect transistor (FET) devices in which the critical source–drain regions are defined by metallic NiSi nanowire sections on p-type Si nanowire. Current versus source–drain voltage (V_{sd}) data (Fig. 4a) are linear to $|V_{\text{sd}}| \leq 1$ V, which suggests that the NiSi/Si contacts behave for practical purposes as ohmic contacts at room temperature, although preliminary temperature-dependent measurements show a barrier at low temperature. The room-temperature behaviour can be explained by the reported segregation of dopant to the NiSi/Si interface during NiSi formation⁹, although other factors may also contribute to the ohmic response at room temperature. Notably, I – V_{sd} data recorded at different back gate

voltages (V_g) exhibit the behaviour expected¹⁷ of a depletion mode p-FET with a high hole mobility of 325 $\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$ (see Methods).

To demonstrate that the active channel of this device was defined by the separation between NiSi nanowire regions, and not the much larger lithographically defined metal contacts, we carried out scanning gate microscopy (SGM) measurements. SGM images recorded with the scanning gate voltages of -9 V (Fig. 4c) and $+9$ V (Fig. 4d) respectively show enhanced (accumulation) and reduced (depletion) conductance only in the silicon region over the overall device (Fig. 4b). These data thus confirm that our approach yields spatially and electronically well-defined metal–semiconductor devices.

Lastly, we have investigated scaling of our approach by assembling¹⁸ crossed-nanowire arrays in which the crossed nanowires function as masks defining active channels in NiSi/Si/NiSi heterostructures (see Methods) and could also function as local gates (Fig. 5a). TEM images recorded on a device following removal of the crossed-nanowire mask (Fig. 5b) show a well-defined silicon channel of 20 nm in this 10-nm-diameter NiSi/Si/NiSi heterostructure. This channel length is comparable to the best that can be achieved in state-of-the-art planar devices¹⁹. The TEM results indicate lateral diffusion of several nanometres during the formation of NiSi, and suggest that it should be possible to prepare shorter channel devices in a well-defined manner simply by varying the diameter of the nanowire mask. More generally, the capability of transforming Si to NiSi in a spatially well-defined manner to form NiSi/Si nanowire heterostructures and superlattices with atomically sharp metal–semiconductor interfaces opens up the possibility of integrating both active devices and high-performance interconnects from a single nanoscale building block. By extending our approach to crossed nanowires as shown above it should become possible to assemble large and dense arrays, perhaps using Langmuir–Blodgett assembly techniques²⁰, of transistors and other devices that could enable hybrid integrated circuits and could represent a key step towards stand-alone integrated nanosystems. $\square$

Methods

Preparation of single-crystal metallic NiSi nanowires

Si nanowires were synthesized via chemical vapour deposition using monodisperse gold nanoclusters (Ted Pella) as catalysts, silane (SiH_4) as the vapour-phase reactant, diborane as the dopant and hydrogen as the carrier gas^{10,11}. The silicon nanowires produced in this manner have clean surfaces with no visible amorphous oxide¹¹. Immediately following synthesis, the growth substrate with uniform diameter, free-standing Si nanowires was loaded into a metal deposition system, and then Ni metal was deposited to a thickness comparable to the average Si nanowire diameter. NiSi nanowires were produced by annealing the Ni-metal-coated Si nanowires at 550 $^\circ\text{C}$. Excess Ni was completely removed by etching (TFG, Transene) for one hour at 50 $^\circ\text{C}$, followed by post-annealing at 600 $^\circ\text{C}$. Each of the annealing steps was carried out in forming gas (N_2 : H_2 ratio, 90:10) for 5 min in a rapid thermal annealer (Heatpulse 610, Metron Technology).

Preparation of NiSi/Si nanowire heterostructures

Si nanowires dispersed in ethanol were deposited on a Si wafer with 600 nm of thermal oxide, and then the substrate was coated with photoresist (Shipley 1813, Rohm and Haas Electronics Materials). The photoresist was exposed for about 2 s on an ABM photoaligner using a simple striped pattern with a 2- μm pitch: 1 μm linewidth and 1 μm spacing. After developing for 1 min, the wafer was transferred to a thermal evaporator and Ni was evaporated with a thickness equal to the average nanowire diameter. After lift-off, the samples were annealed and etched as described above. Ultrasmall NiSi/Si/NiSi nanowire heterostructures were fabricated using crossed Si/SiO₂ core-shell nanowires²¹ as masks to define the lengths of the unreacted Si regions. The crossed-nanowire structures were assembled on Si_3N_4 membrane window grids (Structure Probe) by fluidic assembly¹⁸, and then Ni was evaporated and annealed as described above. The Si/SiO₂ core-shell nanowires were removed with hydrogen fluoride solution (Transene) to enable direct TEM imaging of the NiSi/Si/NiSi heterostructure on the Si_3N_4 membranes.

Calculation of device characteristics

The hole mobility, μ , is computed using a standard model¹⁷. In the small bias linear transport region, the mobility is expressed in terms of the transconductance, g_m , as $g_m = \mu C V_{\text{sd}} / L^2$, where V_{sd} is the source–drain voltage, L is the device channel length, and C is the gate capacitance, which was estimated using $C = 2\pi\epsilon_0 L / \ln(2t_{\text{ox}}/r)$ (where ϵ is the effective dielectric constant, t_{ox} is the SiO₂ dielectric thickness, and r is the nanowire radius). For the device in Fig. 3a, g_m was evaluated for V_g from -3 to $+3$ V at $V_{\text{sd}} = -1$ V, and has a value of 275 nS .

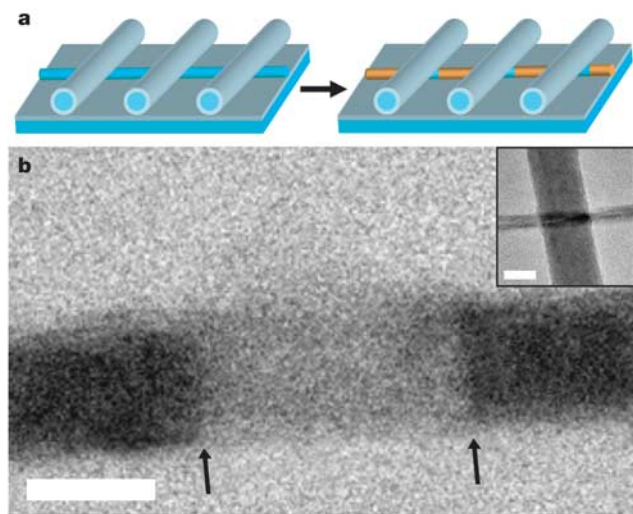


Figure 5 Self-aligned nanoscale NiSi/Si/NiSi devices. **a**, Schematic illustrating Si nanowire (blue) crossed with three Si/SiO₂ core (blue)–shell (grey) nanowires²¹. Deposition, annealing and removal of excess Ni yields NiSi (brown) regions separated by Si in the nanowire. **b**, TEM image of the NiSi/Si/NiSi nanowire heterostructure. The dark regions correspond to NiSi and the light region to Si with NiSi/Si interfaces highlighted by black arrows. Scale bar is 10 nm. Inset, TEM image of the same nanowire before silicidation. The crossed Si/SiO₂ core–shell nanowire (approximately vertical in image) was used as a mask to define the Si region and removed after silicidation. Scale bar is 20 nm. The sample was prepared and imaged on a 50-nm-thick Si_3N_4 membrane.

Received 1 March; accepted 24 May 2004; doi:10.1038/nature02674.

1. Yao, Z., Dekker, C. & Avouris, Ph. Electrical transport through single-wall carbon nanotubes. *Top. Appl. Phys.* **80**, 147–171 (2001).
2. McEuen, P. L., Fuhrer, M. S. & Park, H. Single-walled carbon nanotube electronics. *IEEE Trans. Nanotechnol.* **1**, 78–85 (2002).
3. Dai, H. Carbon nanotubes: synthesis, integration, and properties. *Acc. Chem. Res.* **35**, 1035–1044 (2002).
4. Lieber, C. M. Nanoscale science and technology: Building a big future from small things. *MRS Bull.* **28**, 486–491 (2003).
5. Lieber, C. M. The incredible shrinking circuit. *Sci. Am.* **285**, 58–64 (2001).
6. Cui, Y., Zhong, Z., Wang, D., Wang, W. U. & Lieber, C. M. High performance silicon nanowire field effect transistors. *Nano Lett.* **3**, 149–152 (2003).
7. Javey, A., Guo, J., Wang, Q., Lundstrom, M. & Dai, H. Ballistic carbon nanotube field-effect transistors. *Nature* **424**, 654–657 (2003).
8. Heinze, S. *et al.* Carbon nanotubes as Schottky barrier transistors. *Phys. Rev. Lett.* **89**, 106801 (2002).
9. Morimoto, T. *et al.* Self-aligned nickel-mono-silicide technology for high-speed deep submicrometer logic CMOS ULSI. *IEEE Trans. Electron Devices* **42**, 915–922 (1995).
10. Cui, Y., Lauhon, L. J., Gudiksen, M. S., Wang, J. & Lieber, C. M. Diameter-controlled synthesis of single-crystal silicon nanowires. *Appl. Phys. Lett.* **78**, 2214–2216 (2001).
11. Wu, Y. *et al.* Controlled growth and structures of molecular-scale silicon nanowires. *Nano Lett.* **4**, 433–436 (2004).
12. Toman, K. The structure of NiSi. *Acta Crystallogr.* **4**, 462–464 (1951).
13. Meyer, B. *et al.* Intrinsic properties of NiSi. *J. Alloys Compounds* **262/263**, 235–237 (1997).
14. Pierce, D. G. & Brusius, P. G. Electromigration: a review. *Microelectron. Reliab.* **37**, 1053–1072 (1997).
15. Yao, Z., Kane, C. L. & Dekker, C. High-field electrical transport in single-wall carbon nanotubes. *Phys. Rev. Lett.* **84**, 2941–2944 (2000).
16. *International Technology Roadmap for Semiconductors 2003* edn (<http://public.itrs.net/Files/2003ITRS/Interconnect2003.pdf>) (2003).
17. Sze, S. M. *Physics of Semiconductor Devices* 438–445 (John Wiley & Sons, New York, 1981).
18. Huang, Y., Duan, X., Wei, Q. & Lieber, C. M. Directed assembly of one-dimensional nanostructures into functional networks. *Science* **291**, 630–633 (2001).
19. Chau, R. *et al.* Silicon nano-transistors for logic applications. *Physica E* **19**, 1–5 (2003).
20. Whang, D., Jin, S., Wu, Y. & Lieber, C. M. Large-scale hierarchical organization of nanowire arrays for integrated nanosystems. *Nano Lett.* **3**, 1255–1259 (2003).
21. Lauhon, L. J., Gudiksen, M. S., Wang, D. & Lieber, C. M. Epitaxial core-shell and core-multishell nanowire heterostructures. *Nature* **420**, 57–61 (2002).

Acknowledgements We thank M. C. McAlpine, C. J. Barrelet and D. C. Bell for discussions. C.M.L. thanks the Defense Advanced Research Projects Agency and Intel for support of this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.M.L. (cml@cmliris.harvard.edu).

Evidence for deep-water production in the North Pacific Ocean during the early Cenozoic warm interval

Deborah J. Thomas

Department of Oceanography, Texas A&M University, College Station, Texas 77843-3146, USA

The deep-ocean circulation is responsible for a significant component of global heat transport. In the present mode of circulation, deep waters form in the North Atlantic and Southern oceans where surface water becomes sufficiently cold and dense to sink. Polar temperatures during the warmest climatic interval of the Cenozoic era (~65 to 40 million years (Myr) ago) were significantly warmer than today, and this may have been a consequence of enhanced oceanic heat transport¹. However, understanding the relationship between deep-ocean circulation and ancient climate is complicated by differences in oceanic gateways², which affect where deep waters form and how they circulate. Here I report records of neodymium isotopes from two cores in the Pacific Ocean that indicate a shift in deep-water production from the Southern Ocean to the North Pacific ~65 Myr ago. The source of deep waters reverted back to the Southern Ocean 40 Myr ago. The

relative timing of changes in the neodymium and oxygen isotope records indicates that changes in Cenozoic deep-water circulation patterns were the consequence, not the cause, of extreme Cenozoic warmth.

Slow deep-water renewal in the modern North Pacific is a consequence of climatic and tectonic factors. Relatively high precipitation in the North Pacific³ results in sea-surface salinities too low to enable downwelling (~33 parts per thousand (p.p.t.) as compared to ~35 p.p.t. in the North Atlantic³). Further, the Bering Strait sill does not permit dense, cold Arctic bottom waters to enter the North Pacific⁴. In addition, most of the bottom waters formed in the Ross Sea (the Pacific sector of the Southern Ocean) are prevented from flowing northward into the Pacific basin by the eastward flow of the Antarctic circumpolar current and the mid-ocean ridge⁴.

However, a different set of climatic and tectonic boundary conditions prevailed during the early Cenozoic interval of extreme warmth, about 65–40 Myr ago. Global deep-water temperatures were about 8–12 °C (ref. 5), too warm for permanent polar ice to exist. The configuration of continents and ocean basins was also different from the present (Fig. 1). The Tasman Sea and Drake Passage had not yet opened, and Tethyan and Caribbean seaways still existed². In addition, the Norwegian and Greenland seas began forming ~55 Myr ago⁶, so North Atlantic deep waters would not have existed during this time interval.

Such different thermal and palaeogeographic regimes should have affected deep-sea circulation patterns, and conversely, a different mode of thermohaline circulation may have influenced the evolution of global climate. To investigate the relationship between early Cenozoic climate and thermohaline circulation patterns, I generated records of the neodymium isotopic composition of fossil fish debris from Ocean Drilling Program (ODP) Sites 1209 and 1211 (Shatsky rise, in the present-day northwestern Pacific).

Neodymium isotopic ratios (¹⁴³Nd/¹⁴⁴Nd, expressed as ϵ_{Nd} ; ref. 7) track deep-water mass circulation because Nd has a short oceanic residence time (~1,000 yr; ref. 8) relative to oceanic mixing (~1,500 yr; ref. 9). The ϵ_{Nd} composition of individual deep-water masses is derived from the composition of Nd discharging into the source regions of the water masses^{10–12}. South Pacific surface

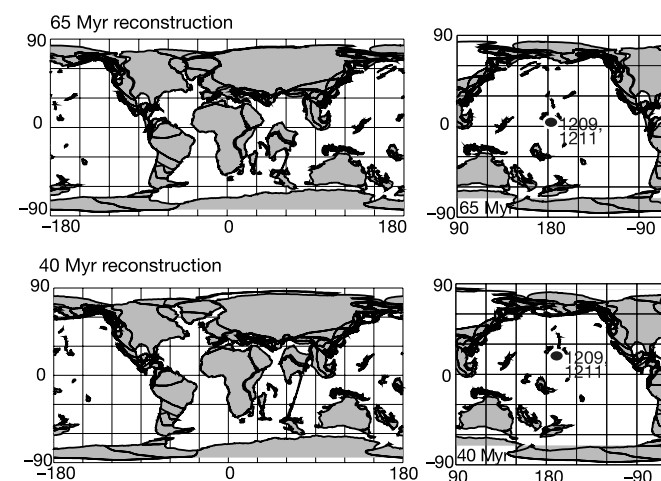


Figure 1 Palaeogeographic reconstructions during two intervals of the early Cenozoic. The location of Ocean Drilling Program (ODP) Sites 1209 and 1211 are shown for 65 Myr ago (top panels) and 40 Myr ago (bottom panels). Note that the location of Sites 1209 and 1211 did not change significantly over the time interval investigated. Reconstructions created by the Ocean Drilling Stratigraphic Network (www.odsnet.org).

water ϵ_{Nd} values are highly radiogenic (~ 0) but are underlain by the non-radiogenic ϵ_{Nd} values (about 8) of northward flowing Antarctic bottom waters (AABW)¹³. North Pacific ϵ_{Nd} depth profiles are much less stratified despite the existence of distinct intermediate-, deep- and bottom-water masses¹⁴, a consequence of slow deep-water renewal in this region. Vertical exchange of Nd between radiogenic Pacific Intermediate Water and underlying AABW ($\epsilon_{\text{Nd}} \approx -8$) generates the relatively radiogenic ϵ_{Nd} signature (about 5) of the deep North Pacific.

To reconstruct ancient Pacific deep-sea circulation, I analysed the Nd isotopic composition of teeth and bones from fossil fish. Fish debris record the ϵ_{Nd} signal of oceanic deep and bottom waters at the sediment surface^{15,16}, and this signal is resistant to diagenetic alteration^{17,18}. The advantages of fossil fish debris over other benthic sedimentary phases (for example, Fe-Mn crusts, benthic foraminifera) are their widespread geographic and stratigraphic distribution, and their resistance to dissolution in corrosive bottom waters.

Fish debris Nd isotope data from ODP Leg 198 Sites 1209 and

Table 1 Shatsky rise Nd isotope data

Depth (m.b.s.f.)	Age (Myr)	$^{143}\text{Nd}/^{144}\text{Nd}$	$\epsilon_{\text{Nd}}(t)$
Site 1209			
114.60	33.88	0.512414 ± 10	-4.52 ± 0.20
133.58	38.30	0.512400 ± 07	-4.60 ± 0.14
141.61	41.28	0.512427 ± 12	-4.00 ± 0.24
146.89	45.02	0.512442 ± 09	-3.61 ± 0.18
146.89	45.02	0.512444 ± 10	-3.57 ± 0.20
152.62	45.99	0.512476 ± 09	-2.93 ± 0.18
163.60	48.07	0.512445 ± 08	-3.48 ± 0.16
167.39	48.62	0.512462 ± 09	-3.13 ± 0.18
171.60	49.47	0.512448 ± 07	-3.38 ± 0.14
177.60	50.35	0.512465 ± 09	-3.03 ± 0.18
185.60	53.13	0.512443 ± 12	-3.39 ± 0.24
204.60	56.68	0.512456 ± 10	-3.05 ± 0.20
210.21	57.68	0.512453 ± 08	-3.07 ± 0.16
213.21	58.51	0.512449 ± 09	-3.14 ± 0.18
214.10	58.75	0.512451 ± 08	-3.10 ± 0.16
214.66	58.91	0.512453 ± 08	-3.06 ± 0.16
223.59	61.16	0.512434 ± 07	-3.37 ± 0.14
226.60	61.89	0.512418 ± 11	-3.65 ± 0.22
230.11	62.92	0.512459 ± 09	-2.84 ± 0.18
236.11	64.88	0.512438 ± 08	-3.20 ± 0.16
263.71	68.66	0.512350 ± 08	-4.82 ± 0.16
265.83	68.96	0.512356 ± 07	-4.69 ± 0.14
268.59	69.35	0.512438 ± 11	-3.09 ± 0.22
269.20	69.44	0.512374 ± 09	-4.32 ± 0.18
270.09	69.57	0.512423 ± 08	-3.37 ± 0.16
Site 1211			
83.70	35.80	0.512372 ± 08	-5.22 ± 0.16
86.70	39.48	0.512388 ± 06	-4.81 ± 0.12
88.70	40.84	0.512404 ± 08	-4.46 ± 0.16
97.70	49.41	0.512475 ± 10	-2.86 ± 0.20
101.20	50.30	0.512440 ± 08	-3.52 ± 0.16
107.20	51.45	0.512428 ± 08	-3.72 ± 0.16
110.70	52.28	0.512472 ± 10	-2.85 ± 0.20
116.55	54.81	0.512424 ± 09	-3.71 ± 0.18
117.21	55.87	0.512433 ± 12	-3.52 ± 0.24
123.20	57.61	0.512422 ± 05	-3.69 ± 0.10
128.20	63.05	0.512431 ± 09	-3.38 ± 0.18
137.70	66.05	0.512391 ± 07	-4.08 ± 0.14
147.20	67.77	0.512388 ± 07	-4.10 ± 0.15
156.70	69.49	0.512369 ± 08	-4.42 ± 0.16

Fish debris was hand picked from the $>63\text{-}\mu\text{m}$ size fraction of washed samples, then cleaned using an established reductive/oxidative cleaning protocol^{25,26}. Samples were analysed as NdO^+ using a multi-collector Micromass Sector 54 at UNC-CH. Monitor peak ($^{144}\text{Nd}^{16}\text{O}$) beams of $\sim 1\text{V}$ were achieved by introducing pure oxygen into the source via a leak valve. External analytical precision based upon replicate analysis of the UNC Ames Nd standard (as NdO^+) was 0.512140 ± 0.000014 (2σ), and analysis of the international standard JNd, (ref. 27) yielded 0.512111 ± 0.000022 , which is calibrated relative to the La Jolla standard (0.511858) as 0.512116 . Reported errors are within-run 2σ values, which corresponds to a minimum uncertainty of $\pm 0.28\epsilon$ units when combined with the external precision. All reported errors are in the last two decimal places. The procedural blank is $\sim 50\text{pg}$ and is considered negligible. Nd isotope values are reported using the epsilon notation, ϵ_{Nd} , which normalizes the analysed $^{143}\text{Nd}/^{144}\text{Nd}$ ratio to the present-day bulk Earth value of CHUR (chondritic uniform reservoir), 0.512638 (ref. 9). The Sm isotopic composition of five samples yielded a mean $^{147}\text{Sm}/^{144}\text{Nd}$ ratio of 0.131 from which $\epsilon_{\text{Nd}}(t)$ values were calculated. $\epsilon_{\text{Nd}}(t)$ represents the isotopic ratio corrected for the evolution of the sample and CHUR through time as ^{147}Sm constantly decays. Ages were obtained by linearly interpolating sedimentation rates between biostratigraphic datums determined during ODP Leg 198 (ref. 28). m.b.s.f., metres below sea floor.

1211 (Table 1) demonstrate the evolution of deep-water mass composition over the period ~ 70 to 30 Myr ago. The trend in ϵ_{Nd} values at ODP Site 1209 (palaeodepth $\sim 2,300\text{m}$) begins with an increase from 4.8 to 3.2 over the interval $68.66\text{--}64.88$ Myr ago (Fig. 2). Over the next ~ 20 Myr, ϵ_{Nd} values fluctuate by up to 0.5ϵ units about an average value of 3.2 . Then ϵ_{Nd} values decrease from 2.9 to 4.5 over the interval $45.99\text{--}33.89$ Myr ago. The parallel ϵ_{Nd} record generated from nearby ODP Site 1211 (palaeodepth $\sim 2,900\text{m}$) is characterized by an increase in ϵ_{Nd} values from 4.4 (69.49 Myr ago) to 3.4 (63.05 Myr ago) (Fig. 2). From 63.05 to 49.41 Myr ago, Site 1211 ϵ_{Nd} values remain relatively high with a mean of 3.4 , although a series of $\sim 0.8\epsilon$ unit oscillations occurs from 54.81 to 49.41 Myr ago. After 49.41 Myr ago, ϵ_{Nd} values decrease to 5.2 by 35.80 Myr ago.

The similarity in the overall ϵ_{Nd} trends at both Shatsky rise sites indicates that these sites were bathed by a common deep-water mass that spanned a depth range of at least $2,300$ to $2,900\text{m}$. Although the timing and magnitude of short-term ϵ_{Nd} oscillations at both sites do not coincide (probably attributable to variations in sampling resolution and stratigraphic completeness), the long-term trends recorded at both sites indicate a period of major regional oceanographic change during the early Palaeogene. Deep waters in the central subtropical Pacific evolved from a more non-radiogenic composition (about 4.5 to 5) from the latest Cretaceous to the Early Palaeocene (~ 70 to 64 Myr ago), then remained relatively radiogenic (about 3 to 3.5) for the next ~ 20 Myr, and subsequently became more non-radiogenic from ~ 45 to 33 Myr ago.

The shifts of about 1.5ϵ units recorded at both sites reflect a fundamental change in the source of deep waters bathing Shatsky rise. The alternative, that motion of the Pacific plate positioned Shatsky rise within a different deep-water mass, is ruled out on the basis of negligible northward plate migration over the interval $\sim 65\text{--}40$ Myr ago (Fig. 1), as well as the fact that deep-water Nd isotopic values shifted back towards more radiogenic values by ~ 40 Myr ago. To place the magnitude of this long-term ϵ_{Nd} shift in palaeoceanographic perspective, glacial to interglacial changes in Atlantic Ocean thermohaline circulation patterns resulted in ϵ_{Nd} changes of ~ 1 to 2ϵ units¹⁹.

The Site 1209 and 1211 ϵ_{Nd} data contain evidence for radiogenic and non-radiogenic (albeit radiogenic when compared to Southern

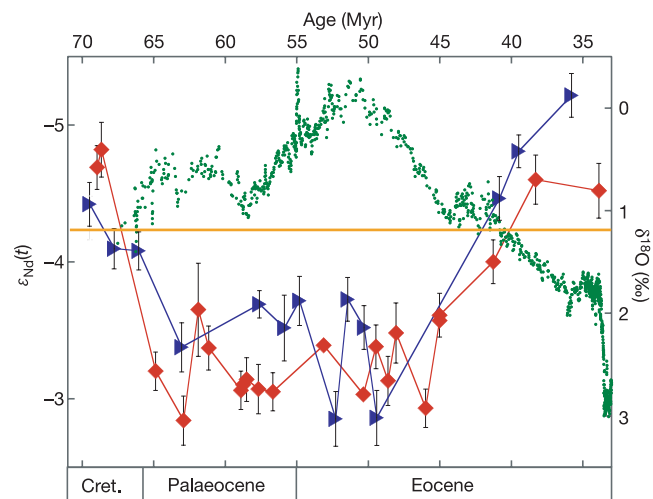


Figure 2 Neodymium isotope records from ODP Sites 1209 and 1211. Nd isotope data: red diamonds, ODP Site 1209; blue triangles, Site 1211. Error bars represent two standard deviations. Benthic foraminiferal oxygen isotope values (green dots)⁵ provide an estimate of deep-sea temperature. The orange line represents a deep-sea temperature of $\sim 7^\circ\text{C}$ calculated from the oxygen isotope values⁵.

Ocean and Atlantic values) end-member sources of deep water. Radiogenic Nd is supplied to the modern Pacific via weathering and drainage of the circum-Pacific arc volcanics¹⁰. The tectonic process that generated the arc volcanics, namely the subduction of the Kula and Farallon plates, had been underway at least since ~110 Myr ago²⁰. Thus, the North Pacific was probably the source of radiogenic Nd to Late Cretaceous and early Palaeogene oceans. The shift of the relatively radiogenic deep-water Nd isotopic composition recorded at Shatsky rise ($\epsilon_{\text{Nd}} \approx 5$) towards more radiogenic values at ~66 Myr ago (about 3.5) implies a more focused transfer of highly radiogenic surface waters to the deep ocean, at least to a depth of 2,900 m. This transfer of radiogenic Nd into the deep Pacific is best explained by deep-water production in the North Pacific.

The non-radiogenic ϵ_{Nd} end-member signature was produced in the Southern Ocean during the early Palaeogene²¹. There are insufficient data to determine in which sector of the Southern Ocean these deep waters formed, but this non-radiogenic end-member composition is still clearly distinct from deep waters with a radiogenic North Pacific signature. Formation of deep waters in the North Pacific represents a mode of thermohaline circulation considerably different from the modern, a mode that may have been dictated by the tectonic and climatic boundary conditions of the early Cenozoic.

Comparison of the Shatsky rise ϵ_{Nd} records to the Cenozoic record of global deep-sea temperatures derived from oxygen isotope analyses of benthic foraminifera⁵ (Fig. 2) indicates that major shifts in deep-water ϵ_{Nd} composition are bracketed by the warmest deep-sea conditions of the Cenozoic. The change from non-radiogenic, Southern Ocean ϵ_{Nd} values to radiogenic, North Pacific ϵ_{Nd} values coincides approximately with warming following the Cretaceous/Tertiary boundary event. Shatsky rise deep waters carried a North Pacific deep-water ϵ_{Nd} signal throughout the warmest portion of the Cenozoic record, when global deep-water temperatures were above ~7°C (Fig. 2). As global deep waters continued to cool during the Eocene, the source of deep waters shifted back to the Southern Ocean.

The relative timing of the switch from southern, high-latitude deep-water production to the Pacific northern high latitudes during the period of general global warming—and the subsequent reversion back to a Southern Ocean source of deep waters during the protracted Eocene cooling trend—implies that the long-term evolution of thermohaline circulation patterns was a consequence and not a cause of global climate change. Although the resolution of the Nd and oxygen isotopic records is coarse over the interval ~69–65 Myr ago, the onset of the Nd isotopic change, particularly evident in the Site 1211 record, lags the decrease (warming) in oxygen isotope values (Fig. 2). However, the switch back to a Southern Ocean deep-water source region clearly occurs in the midst of global cooling. Thus the global cooling that marked the end of the extreme early Cenozoic warmth also ended deep-water production in the North Pacific.

Model simulations that account for continental runoff and precipitation–evaporation distributions²² predict downwelling in the North Pacific under early Cenozoic tectonic and climatic boundary conditions. The transition to extreme global warmth during the early Cenozoic must have altered the distribution of precipitation patterns, rendering Southern Ocean deep-water formation regions slightly fresher, and/or North Pacific deep-water formation regions slightly more saline and hence sufficiently dense to permit significant downwelling of North Pacific waters.

The implications of long-term climate change forcing the evolution of thermohaline circulation patterns are significant. A mechanism of enhanced poleward heat via the deep oceans can no longer be invoked as a cause of extreme global warmth during the early Cenozoic (and potentially the mid to late Cretaceous warm interval), as deep-water circulation now has been shown to respond to global climate change on long timescales. In addition, the change to

deep-water production in the North Pacific could not have caused the global deep-water warming. Trends exhibited by the global compilation of benthic foraminiferal oxygen isotope data⁵ (Fig. 2) indicate that deep-water temperatures warmed (and cooled) in each basin at approximately the same time. Thus, a switch in deep-water source regions would not necessarily affect deep-water temperatures if source regions warm (or cool) under global control, particularly on longer timescales. Alternative mechanisms for general global warming and warming of global deep-sea temperatures during the early Cenozoic need to be further explored. The most likely candidates are those that involve increased atmospheric inventories of carbon dioxide, such as effusive volcanism in the North Atlantic igneous province²³, or widespread burning of terrestrial peat deposits²⁴. □

Received 25 February; accepted 5 May 2004; doi:10.1038/nature02639.

- Sloan, L. C. & Rea, D. K. Atmospheric carbon dioxide and early Eocene climate: A general circulation modeling sensitivity study. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **119**, 275–292 (1995).
- Frakes, L. A. & Kemp, E. L. Influence of continental positions on early Tertiary climates. *Nature* **240**, 97–100 (1972).
- Levitus, S. *Climatological Atlas of the World Ocean* (Professional Paper 13, NOAA, Washington DC, 1982).
- Tomczak, M. & Godfrey, J. S. *Regional Oceanography: An Introduction* (Pergamon, London, 1994).
- Zachos, J. C., Pagani, M., Sloan, L. C., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
- Saunders, A. D., Fitton, J. G., Kerr, A. C., Norry, M. J. & Kent, R. W. In *Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism* (eds Mahoney, J. J. & Coffin, M. F.) 45–93 (AGU Geophys. Monogr. 100, American Geophysical Union, Washington DC, 1997).
- DePaolo, D. J. & Wasserburg, G. J. Nd isotopic variations and petrogenetic models. *Geophys. Res. Lett.* **3**, 248–252 (1976).
- Tachikawa, K., Jeandel, C. & Roy-Barman, M. A new approach to the Nd residence time in the ocean: the role of atmospheric inputs. *Earth Planet. Sci. Lett.* **170**, 433–446 (1999).
- Broecker, W. S., Gerard, R., Ewing, M. & Heezen, B. C. Natural radiocarbon in the Atlantic Ocean. *J. Geophys. Res.* **65**, 2903–2931 (1960).
- Goldstein, S. L. & Jacobsen, S. B. Nd and Sr isotope systematics of river water suspended material: Implications for crustal evolution. *Earth Planet. Sci. Lett.* **87**, 249–265 (1987).
- Halliday, A. N., Davidson, J. P., Holden, P., Owen, R. M. & Olivarez, A. M. Metalliferous sediments and the scavenging residence time of Nd near hydrothermal vents. *Geophys. Res. Lett.* **19**, 761–764 (1992).
- Jones, C. E., Halliday, A. N., Rea, D. K. & Owen, R. M. Neodymium isotopic variations in the North Pacific modern silicate sediment and the insignificance of detrital REE contributions to seawater. *Earth Planet. Sci. Lett.* **127**, 55–66 (1994).
- Piegras, D. J. & Jacobsen, S. B. The isotopic composition of neodymium in the North Pacific. *Geochim. Cosmochim. Acta* **52**, 1373–1381 (1988).
- Talley, L. D. Distribution and formation of North Pacific intermediate water. *J. Phys. Oceanogr.* **23**, 517–537 (1993).
- Reynard, B., Lecuyer, C. & Grandjean, P. Crystal-chemical controls on rare-earth element concentrations in fossil biogenic apatites and implications for paleoenvironmental reconstructions. *Chem. Geol.* **155**, 233–241 (1999).
- Staudigel, H., Doyle, P. & Zindler, A. Sr and Nd isotope systematics in fish teeth. *Earth Planet. Sci. Lett.* **76**, 45–56 (1985).
- Martin, E. E. & Scher, H. D. Preservation of seawater Sr and Nd isotopes in fossil fish teeth: bad news and good news. *Earth Planet. Sci. Lett.* **220**, 25–39 (2004).
- Thomas, D. J. In *Isotopic and Elemental Tracers of Late Cenozoic Climate Change* (eds Surge, D. & Mora, G.) (GSA Spec. Publ., Geological Society of America, in the press).
- Rutberg, R. L., Hemming, S. R. & Goldstein, S. L. Reduced North Atlantic deep water flux to the glacial Southern Ocean inferred from neodymium isotope ratios. *Nature* **405**, 935–938 (2000).
- Larson, R. L. & Pitman, W. C. III Worldwide correlation of Mesozoic magnetic anomalies and its implications. *Geol. Soc. Am. Bull.* **83**, 3645–3661 (1972).
- Thomas, D. J., Bralower, T. J. & Jones, C. E. Neodymium isotopic reconstruction of late Paleocene–early Eocene thermohaline circulation. *Earth Planet. Sci. Lett.* **209**, 309–322 (2003).
- Bice, K. L., Barron, E. J. & Peterson, W. H. Continental runoff and early Cenozoic bottom-water sources. *Geology* **25**, 951–954 (1997).
- Eldholm, O. & Thomas, E. Environmental impact of volcanic margin formation. *Earth Planet. Sci. Lett.* **117**, 319–329 (1993).
- Kurtz, A. C., Kump, L. R., Arthur, M. A. & Paytan, A. Early Cenozoic decoupling of the global carbon and sulfur cycles. *Paleoceanography* **18**, 14–14–14 (2003).
- Boyle, E. A. Cadmium, zinc, copper, and barium in foraminifera tests. *Earth Planet. Sci. Lett.* **53**, 11–35 (1981).
- Boyle, E. A. & Keigwin, L. D. Comparison of Atlantic and Pacific paleochemical records for the last 250,000 years: changes in deep ocean circulation and chemical inventories. *Earth Planet. Sci. Lett.* **76**, 135–150 (1985).
- Tanaka, T. et al. Jndi-1: a neodymium isotopic reference in consistency with La Jolla neodymium. *Chem. Geol.* **168**, 279–281 (2000).
- Bralower, T. J. et al. *Proc. ODP Init. Rep.* [online] **198**, (http://www-odp.tamu.edu/publications/198_IR/chap_05/chap_05.htm) (http://www-odp.tamu.edu/publications/198_IR/chap_07/chap_07.htm) (2002).

Acknowledgements The author thanks T. Bralower, U. Rohl and N. Slowey for discussions, the science party, staff and crew of ODP Leg 198, and the Ocean Drilling Program for supplying sample material. This work was supported by JOI.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to the author (dthomas@ocean.tamu.edu).

Role of metal-reducing bacteria in arsenic release from Bengal delta sediments

Farhana S. Islam¹, Andrew G. Gault¹, Christopher Boothman¹, David A. Polya¹, John M. Charnock^{1,2}, Debashis Chatterjee³ & Jonathan R. Lloyd¹

¹Department of Earth Sciences and Williamson Research Centre for Molecular Environmental Science, The University of Manchester, Manchester M13 9PL, UK

²CCLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, UK

³Department of Chemistry, The University of Kalyani, West Bengal, 731 235, India

The contamination of ground waters, abstracted for drinking and irrigation, by sediment-derived arsenic threatens the health of tens of millions of people worldwide, most notably in Bangladesh and West Bengal^{1–3}. Despite the calamitous effects on human health arising from the extensive use of arsenic-enriched ground waters in these regions, the mechanisms of arsenic release from sediments remain poorly characterized and are topics of intense international debate^{4–8}. We use a microcosm-based approach to investigate these mechanisms: techniques of microbiology and molecular ecology are used in combination with aqueous and solid phase speciation analysis of arsenic. Here we show that anaerobic metal-reducing bacteria can play a key role in the mobilization of arsenic in sediments collected from a contaminated aquifer in West Bengal. We also show that, for the sediments in this study, arsenic release took place after Fe(III) reduction, rather than occurring simultaneously. Identification of the critical factors controlling the biogeochemical cycling of arsenic is one important contribution to fully informing the development of effective strategies to manage these and other similar arsenic-rich ground waters worldwide.

Potential mechanisms for the release of arsenic into ground water in Bengali shallow alluvial sedimentary aquifers have been the focus of intense debate. The oxidation of arsenic-rich pyrite in aquifer sediments has been proposed as one possible mechanism^{4,5}. Other studies have suggested that the reductive dissolution of arsenic-rich Fe(III) oxyhydroxides deeper in the aquifer may lead to the release of arsenic into the ground water^{2,6,7,9}. Additional factors that may add further complication to potential arsenic-release mechanisms from sediments include the predicted mobilization of sorbed arsenic by phosphate generated from the intensive use of fertilizers¹⁰, by carbonate¹¹ produced via microbial metabolism⁹, or by changes in the sorptive capacity of ferric oxyhydroxides².

Although the biotic and abiotic processes described above may all play a role in arsenic mobilization, it has been concluded that microorganisms play the defining role in catalysing the redox transformations that ultimately control the mobility of the metalloid⁸. It has also been noted recently that “there is an immediate research need for a fuller understanding of the role(s) of subsurface microbes in mobilizing arsenic in aquifers”⁸. In order to answer this important question, we used sediments from a well-characterized¹²

arsenic-rich test site in the Nadia district, West Bengal, to identify the biogeochemical conditions that promote maximal arsenic release from contaminated aquifers in the Ganges delta. Samples were taken from a depth of 13 m where arsenic-rich ground waters ($>40 \mu\text{g l}^{-1}$) have been found¹³ (see Supplementary Information for further data on sediment and groundwater characteristics, and sampling methods). The sediments were mixed with simulated ground water and incubated at 20 °C under a range of biogeochemical conditions (Fig. 1). Our initial experiments focused on the speciation of arsenic in the pore water, and also the reduction of Fe(III), which is a significant electron acceptor in these sediments and can support the growth of organisms capable of respiring using sediment-bound As(v)¹⁴. Arsenic speciation in the aqueous fraction was monitored using a sensitive coupled IC-ICP-MS system¹⁵, while Fe(II) (in the pore water and bound to sediment material) was monitored using a ferrozine based assay after extraction with 0.5 M HCl¹⁶.

Sediments incubated under aerobic conditions showed negligible reduction of Fe(III) with time, or release of arsenic from the sediments (Fig. 1a). Incubation under anaerobic conditions, however, resulted in Fe(III) reduction concomitant with arsenic mobilization (Fig. 1b). Fe(II) concentrations in the sediment fraction increased from approximately 3.5 mmole l^{-1} to 6 mmole l^{-1} after 38 days incubation, concomitant with the release of 13 nM arsenic, principally as As(III), into the pore water. Throughout these experiments, negligible concentrations of soluble Fe(II) were released into the pore water. Addition of acetate as a potential electron donor for metal reduction and a proxy for organic matter¹⁷ resulted in a marked stimulation in the rate of Fe(III) reduction followed by arsenic release (Fig. 1c). Initially, predominantly dissolved As(v) was detected (up to 23 nM at 15 days), followed by

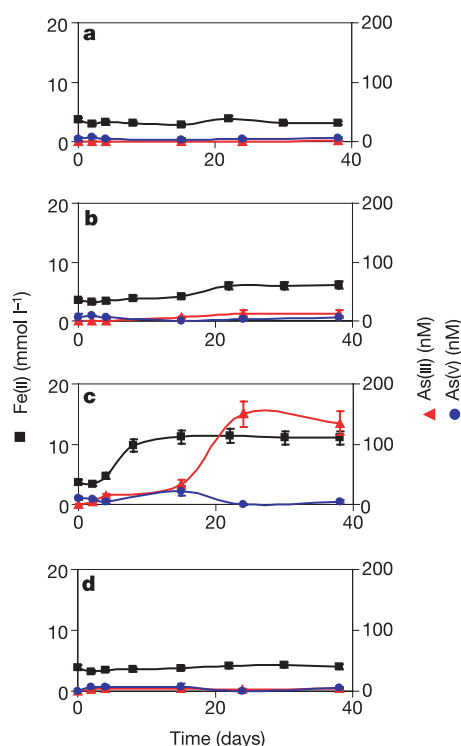


Figure 1 Reduction of Fe(III), and mobilization of arsenic in microcosms containing Bengali sediments incubated under a range of biogeochemical regimes. **a**, Aerobically; **b**, anaerobically; and **c**, anaerobically with 4 g l^{-1} sodium acetate as a proxy for organic matter. ‘Abiotic’ control sediments with added acetate were autoclaved before incubation (**d**). Black squares, Fe(II); red triangles, As(III); purple circles, As(V). Each point and error bar represents the mean and standard deviation of three replicate experiments.

predominately dissolved As(III) which reached concentrations of 150 nM after 24 days. Taking into account the much lower water/sediment ratios in the aquifer (1:8 wt./wt. for a typical porosity of 25% and assuming a sediment density equal to that of quartz) compared to that in our microcosm experiments (2:1 wt./wt.), this As(III) concentration would be equivalent to 2.4 μM or 180 $\mu\text{g L}^{-1}$, broadly comparable with the higher groundwater arsenic concentrations found in this and similar shallow aquifers in Bengal. Increased concentrations of As(III) in pore waters of these sediment experiments is consistent with microbial reduction of As(v). X-ray absorption spectroscopy showed that the predominant (>90%) oxidation state of arsenic in the sediment both before and after metal reduction had been stimulated by acetate was As(v)¹⁸ suggesting that not all of the As(v) was bioavailable for microbial reduction, possibly through its association with recalcitrant crystalline Fe oxides. The observed co-existence of an As(v) dominant solid phase assemblage with an As(III) dominated aqueous phase suggests preferential sorption of As(v) on surfaces of the solid phase assemblage. This is consistent with a number of studies of As sorption on iron oxyhydroxides, (reviewed in ref. 2) although the relative strengths of sorption of As(III) and As(v) is the subject of debate¹⁹. The reduction of Fe(III) and the release of arsenic were clearly decoupled in the sediments that were stimulated with acetate, with Fe(II) concentrations in the sediment fraction approaching the maximum values measured (11 mmol L⁻¹) after only 8 days incubation, while the highest concentrations of As(III) were not recorded until the sediments were incubated for another two weeks (Fig. 1c). That Fe(III) reduction and arsenic release through As(v) reduction are decoupled is not unexpected. Bio-available terminal electron acceptors will be used in sequence governed by the energy yield, derived from the relevant oxidation/reduction potentials. The reduction of Fe(III) followed by reduction of As(v) observed in our experiments can be rationalized in terms of the higher oxidation/reduction potential for Fe(OH)₃/Fe²⁺ at the initially low Fe²⁺ concentrations of our experiments (+241 mV for (Fe²⁺) = 10⁻⁸ M) compared to the oxidation/reduction potential for As(v)/As(III) (+14 mV for equimolar concentrations of As(v) and As(III); calculated from data of refs 20,21; see Supplementary Information).

Decoupling of the reduction of Fe(III) and As(v) may, therefore, reflect adaptation of the respiratory pathways in the microorganisms in the sediments, through dynamic changes in the species that are metabolically active in the microbial community, or through altered expression of the relevant metal reductases in key anaerobes constituting this community. No methylated arsenic species (monomethylarsonic acid or dimethylarsinic acid) were detected (limit of detection, 7 nM as arsenic). Finally, control sediments were autoclaved and then incubated in the presence of acetate. Concentrations of Fe(II), As(v) and As(III) did not increase in these control sediments (Fig. 1d), confirming a role for microorganisms in the reduction of Fe(III) and subsequent mobilization of arsenic.

Having identified incubation conditions that promoted arsenic mobilization, the microbial communities present under the different biogeochemical conditions imposed were analysed, using both cultivation-dependent and molecular (polymerase chain reaction; PCR) techniques.

Initial experiments used a DNA profiling technique that amplified the variable length intergenic spacer region between the genes encoding 16S and 23S rRNA in bacteria (ARISA analysis²²). Using this technique, we noted from ARISA banding patterns (not shown) that there were only minor changes upon a shift from aerobic to anaerobic cultivation conditions, while the addition of acetate resulted in a marked shift in the population, suggesting that microbial metal reduction in our samples was limited by available electron donor. Additions of as little as 1 mM acetate also resulted in a similar shift in the microbial communities as shown by ARISA analysis, with corresponding enhanced levels of As(III) measured in

the pore waters (data not shown). An approximately 500 base pair (bp) region of the 16S rRNA gene was amplified by PCR using broad specificity bacterial primers, cloned and typed using restriction fragment length polymorphism (RFLP) analysis. Of the 46 clones analysed from the starting sediment material, 28 gave distinct banding patterns, suggesting a diverse community of bacteria in the sediments. The majority of these 'RFLP-types' were affiliated with known members of the γ - and β -Proteobacteria (30% and 24% of the clone library respectively; Fig. 2, with a detailed phylogenetic analysis presented as Supplementary Information). In the γ -Proteobacteria, the majority of the clones corresponded to sequences derived from *Pseudomonas* species (26% of total clone library), facultative anaerobes able to respire using alternative electron acceptors including nitrate, but not As(v) or Fe(III)^{8,23}. Other sequences detected that were consistent with anaerobic conditions in the sediments included those related closely to known *Clostridium* species (9% of the clone library), which are involved in the fermentation of organic matter, and can reduce a range of metals²³ including As(v)⁸. Specialist dissimilatory metal-reducing bacteria were also detected; 11% of the clones were affiliated to *Geobacter* type sequences in the δ -Proteobacteria, which have been shown to dominate zones of Fe(III) reduction in the subsurface²⁴ and reduce a wide range of high valence metals with the notable exception of As(v)²³. Finally, it is conceivable that the sample used in this study may have been obtained at the interface between oxic and anoxic regions of the aquifer, as 13% of the clone library aligned with sequences from *Nitrosolobus* species, which are aerobic ammonia-oxidizing bacteria of the β -Proteobacteria²⁵.

When metal (Fe(III) and As(v)) reduction was stimulated by the addition of acetate, there was a marked shift in the population (Fig. 2). The dominant sequences detected in the clone library (70% of the 30 clones analysed) were affiliated with members of the family *Geobacteraceae* in the δ -Proteobacteria. Similar shifts in microbial communities have been noted in other studies in which the reduction of metals was stimulated in the subsurface²⁴. In comparison, *Pseudomonas*-type sequences (γ -Proteobacteria) corresponded to only 3% of the clone library after stimulation with acetate. The marked increase in numbers of sequences of known Fe(III)-reducing bacteria was mirrored in the most probable number (MPN) counts

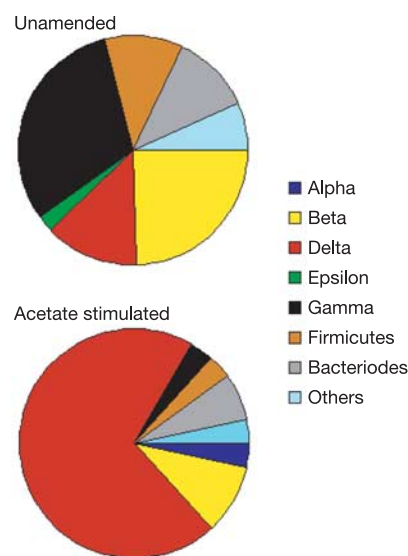


Figure 2 Shifts in the microbial community of the sediment from the Nadia district, West Bengal, after stimulation of anaerobic metal reduction by acetate. 16S rDNA clone libraries were amplified by PCR from untreated sediment (top), and after incubation under anaerobic conditions with added acetate (bottom). Charts show phylogenetic affiliation of the clones. Alpha, beta, delta, epsilon and gamma represent proteobacterial divisions.

on these experiments. Numbers of Fe(III)-reducing bacteria were shown to increase after stimulation with acetate: the following counts per gram sediment were recorded; 55 (aerobic), 220 (anaerobic) and 2.3×10^3 (anaerobic stimulated with acetate). It should be noted that although these counts are satisfactory for comparing the relative amounts of Fe(III)-respiring organisms in the different treatments, MPN counts can underestimate cell numbers by several orders of magnitude.

The organisms that catalyse the reductive mobilization of arsenic remain to be identified irrevocably, and we do not discount the importance of bacterial strains not represented in our clone libraries for this transformation. However, the lack of a coherent phylogenetic grouping or a conserved functional gene for this group of diverse microorganisms⁸ makes it impossible, at present, to search for these organisms using PCR-based techniques, and a novel MPN-based technique for enumerating As(v)-reducing bacteria²⁶ proved irreproducible in our hands. Nevertheless, it remains highly probable that the Fe(III)-reducing bacteria, maintained on the relatively high concentrations of Fe(III) in the sediments, played a major role in the subsequent reduction and release of arsenic from sediments once the bioavailable Fe(III) had been used as an electron acceptor. Indeed, of the 16 species of As(v)-respiring organisms currently in pure culture, none are obligate As(v) reducers⁸. All are 'opportunists' capable of respiring using other electron acceptors, for example Fe(III)⁸. It should be noted that measurements of the metal fractions extracted from our test sediments (before mixing with ground water in our microcosm experiments), using 0.2 M NH_4 oxalate²⁷, showed that they contained 0.7% bioavailable Fe in amorphous and poorly crystalline Fe hydrous oxides, compared to only 1.8 μg As per g.

To confirm whether Fe(III)-reducing bacteria could play a role in the release of arsenic from the West Bengal sediments, a stable enrichment culture of Fe(III)-reducing bacteria was obtained by resubculturing in Fe(III)-containing medium 4 times, washed in artificial ground water and introduced into heat sterilized sediments at a cell concentration similar to those noted in the MPN experiments. The stable culture of Fe(III)-reducing bacteria was indeed able to mobilize arsenic from the sediments, with 94 nM As(III) and 7 nM As(v) noted after 41 days in the pore waters in the presence of added acetate (Fig. 3). Negligible arsenic mobilization was noted in uninoculated sterile controls.

These studies offer direct evidence that anaerobic metal-reducing bacteria play a role in the formation of toxic, mobile As(III) in sediments from the Ganges delta. Our results also suggest that the capacity for arsenic release was severely limited by the availability of

electron donor in the sediments from our test site. Although our experiments are microcosm-based and not a direct simulation of the hydrogeological parameters and organic carbon flux *in situ*, we suggest that our results support theories that the delivery of surface-derived organic carbon²⁸ into subsurface communities (driven, for example, by irrigation pumping⁹) may have a dramatic role in enhancing arsenic mobility in shallow ground waters in the Ganges Delta. Further studies are now warranted on the detailed mechanisms of arsenic release by metal-reducing bacteria isolated from arsenic-contaminated sediments, alongside field-based studies assessing the impact of such transformations *in situ*. □

Methods

Sediment incubations

Sediment was collected from Chakdaha block, Nadia district in West Bengal (GPS location $23^\circ 04' 55'' \text{ N} / 88^\circ 30' 44'' \text{ E}$) using a reverse circulatory drilling method but with samples recovered without the use of drilling fluid to minimize the possibility of contamination with non-indigenous microorganisms. The sample was transported and stored at 4°C under N_2 in a sealed transparent plastic tube to minimize microbial activity before use (see Supplementary Information for a more detailed description of sampling and storage protocols). About 15 g of sediment was mixed with 30 ml artificial ground water based on the constituents of water samples from the study site (MgCl_2 , 0.34 mM; KH_2PO_4 , 0.01 mM; NaHCO_3 , 0.51 mM; K_2CO_3 , 0.025 mM; MgSO_4 , 0.03 mM; KNO_3 , 0.001 mM; and CaCO_3 , 1.85 mM; adjusted to pH 7 using HCl). All glassware was prewashed with 1 M HCl and ultraclean water before use. The bottles were sealed with butyl rubber stoppers, and incubated under aerobic conditions (stoppers pierced with three hypodermic syringe needles), anaerobic conditions (N_2 atmosphere), and anaerobic with added electron donor (4 g l^{-1} acetate). Control experiments comprised autoclaved sediments amended with acetate, incubated under N_2 . Sediments were incubated in the dark at 20°C .

Analytical techniques

Fe(II) was quantified spectrophotometrically after reaction with ferrozine, with total Fe measured after reaction with hydroxylamine¹⁶. Arsenic speciation was analysed by IC-ICP-MS¹⁵. Samples (1 ml) were removed from the bottles in an anaerobic cabinet and passed through a 0.45 μm filter. As(v) and As(III) were separated using a Cetac ANX3206 anion exchange column, housed in a Metrohm 790 Personal IC unit, which was interfaced to a VG PlasmaQuad II ICP-MS. The aqueous mobile phase was 1.7 mM NaHCO_3 / 1.8 mM Na_2CO_3 . Standard reference materials were included in each analytical run; the arsenic concentrations determined were found to agree well with those certified. The limited sample volumes available precluded the determination of other key analytes in the microcosms.

Amplification of 16S rDNA

Sediment samples (0.25 g) were pretreated with 2 ml sodium oxalate solution (0.3 M), passed through a 0.25 μm filter and DNA extracted using a Fast DNA spin kit (UltraClean, Soil DNA Isolation Kit, MO BIO Laboratories). A fragment of the 16S rRNA gene, approximately 520 bp, was amplified by PCR from samples using the broad-specificity primers 8f and 519r²⁴ using an iCycler (BioRad). Purified DNA (5 μl) and 2.5 μl of 25 μM primer stocks were added to the reaction mix to a final volume of 50 μl . The purity of the amplified product was determined by electrophoresis of 10 μl samples in a 1.0% agarose Tris-borate-EDTA (TBE) gel. DNA was stained with ethidium bromide and viewed under short-wave UV light.

RFLP analysis

PCR products were purified using a QIAquick PCR purification kit (Qiagen) and ligated directly into the cloning vector pCR 2.1 (Invitrogen) before transformation into *Escherichia coli* TOP 10 competent cells. White transformants that grew on LB agar containing ampicillin ($100 \mu\text{g ml}^{-1}$) and 40 μl of 40 mg ml^{-1} of X-Gal were screened for an insert using PCR. Primers were complementary to the flanking regions of the PCR insertion site of the pCR 2.1 cloning vector. The PCR method was: initial denaturation at 94°C for 4 min, melting at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 1 min; 35 cycles, with a final extension step at 72°C for 5 min. The PCR products were purified using a QIAquick kit and treated with the restriction endonucleases, *Sau*3A and *Msp*I. The restriction enzyme digests were separated using a 3% metaphor agarose TBE gel.

DNA sequencing and phylogenetic analysis

Nucleotide sequences were determined by the dideoxynucleotide method by cycle sequencing of the purified PCR products. An ABI Prism BigDye Terminator Cycle Sequencing Kit was used in combination with an ABI Prism 877 Integrated Thermal Cycler and ABI Prism 377 DNA Sequencer (Perkin Elmer Applied Biosystems). Sequences (typically 500 bp) were analysed against the NCBI (USA) database using BLAST program packages and matched to known 16S rRNA gene sequences. Gene sequences were aligned using the ClustalX software package and corrected manually. The TREECON package²⁹ was used for distance analysis using the Jukes and Cantor correction and bootstrap resampling (100 times) and a phylogenetic tree constructed from the distance matrix via neighbour joining³⁰.

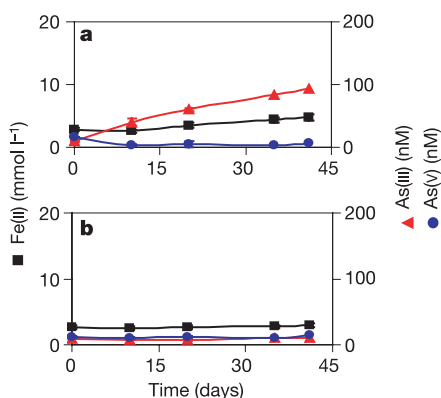


Figure 3 Reduction of Fe(III), and mobilization of arsenic in microcosms containing heat-sterilized Bengali sediments. **a**, Inoculated with an enrichment culture of Fe(III)-reducing bacteria. **b**, Control microcosms were not inoculated with the enrichment culture. Black squares, Fe(II); red triangles, As(III); purple circles, As(v). Each point and error bar represents the mean and standard deviation of three replicate experiments.

Estimation of cell numbers

MPN counting was performed as follows. A tenfold dilution series was prepared from sediment samples, using tubes containing a freshwater medium¹⁶ with 20 mM acetate as the electron donor and 56 mM Fe(III)-citrate as the electron acceptor. The tubes were incubated at 20 °C and scored on a weekly basis for the reduction of Fe(III) using the ferrozine assay¹⁶. The numbers of positive tubes were tabulated and the most probable number of Fe(III)-reducing bacteria estimated by comparison with reference MPN tables.

Received 31 October 2003; accepted 10 May 2004; doi:10.1038/nature02638.

1. Smith, A. H., Lingas, E. O. & Rahman, M. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. *Bull. WHO* **78**, 1093–1103 (2000).
2. Smedley, P. L. & Kinniburgh, D. G. A review of the source, behaviour and distribution of arsenic in natural waters. *Appl. Geochem.* **17**, 517–568 (2002).
3. Chakraborty, D. *et al.* Arsenic calamity in the Indian subcontinent. What lessons have been learnt? *Talanta* **58**, 3–22 (2002).
4. Das, D. *et al.* Arsenic in groundwater in six districts of West Bengal, India. *Environ. Geochem. Health* **18**, 5–15 (1996).
5. Chowdhury, T. R. *et al.* Arsenic poisoning in the Ganges delta. *Nature* **401**, 545–546 (1999).
6. Nickson, R. *et al.* Arsenic poisoning of Bangladesh groundwater. *Nature* **395**, 338 (1998).
7. Nickson, R. T., McArthur, J. M., Ravenscroft, P., Burgess, W. G. & Ahmed, K. M. Mechanism of arsenic release to groundwater, Bangladesh and West Bengal. *Appl. Geochem.* **15**, 403–413 (2000).
8. Oremland, R. S. & Stolz, J. F. The ecology of arsenic. *Science* **300**, 939–944 (2003).
9. Harvey, C. F. *et al.* Arsenic mobility and groundwater extraction in Bangladesh. *Science* **298**, 1602–1606 (2002).
10. Acharyya, S. K. *et al.* Arsenic poisoning in the Ganges delta. *Nature* **401**, 545 (1999).
11. Appelo, C. A. J., Van der Weiden, M. J. J., Tournassat, C. & Charlet, L. Surface complexation of ferrous iron and carbonate on ferrihydrite and the mobilization of arsenic. *Environ. Sci. Technol.* **36**, 3096–3103 (2002).
12. Chatterjee, D. *et al.* Mobilization of arsenic in sedimentary aquifer vis-à-vis subsurface iron reduction processes. *J. Phys. IV France* **107**, 293–296 (2003).
13. Gault, A. G., *et al.* in *Plasma Source Mass Spectrometry: Applications and Emerging Technologies* (eds Holland, J. G. & Tanner, S. D.) 112–126 (Royal Society of Chemistry, Cambridge, UK, 2003).
14. Zobrist, J., Dowdle, P. R., Davis, J. A. & Oremland, R. S. Mobilization of arsenite by dissimilatory reduction of adsorbed arsenate. *Environ. Sci. Technol.* **34**, 4747–4753 (2000).
15. Gault, A. G., Polya, D. A. & Lythgoe, P. R. in *Plasma Source Mass Spectrometry: The New Millennium* (eds Holland, J. G. & Tanner, S. D.) 387–400 (Royal Society of Chemistry, Cambridge, UK, 2001).
16. Lovley, D. R. & Phillips, E. R. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl. Environ. Microbiol.* **54**, 1472–1480 (1988).
17. Lovley, D. R. & Chapelle, F. H. Deep subsurface microbial processes. *Rev. Geophys.* **33**, 365–381 (1995).
18. Gault, A. G. *et al.* Preliminary EXAFS studies of solid phase speciation of arsenic in a West Bengali sediment. *Mineral. Mag.* **67**, 1183–1191 (2003).
19. Dixit, S. & Hering, J. G. Comparison of arsenic(V) and arsenic(III) sorption onto iron oxide minerals: Implications for arsenic mobility. *Environ. Sci. Technol.* **37**, 4182–4189 (2003).
20. Welham, N. J., Malatt, K. A. & Vukcevic, S. The stability of iron phases presently used for disposal from metallurgical systems - a review. *Min. Eng.* **13**, 911–933 (2000).
21. Bard, A. J., Parsons, R. & Jordan, J. *Standard Potentials in Aqueous Solution* (Marcel Dekker, New York, 1985).
22. Ranjard, L. *et al.* Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: Biological and methodological variability. *Appl. Environ. Microbiol.* **67**, 4479–4487 (2001).
23. Lloyd, J. R. Microbial reduction of metals and radionuclides. *FEMS Microbiol. Rev.* **27**, 411–425 (2003).
24. Holmes, D. E., Finneran, K. T. & Lovley, D. R. Enrichment of *Geobacteraceae* associated with stimulation of dissimilatory metal reduction in uranium-contaminated aquifer sediments. *Appl. Environ. Microbiol.* **68**, 2300–2306 (2002).
25. Teske, A., Alm, E. & Regan, J. M. Evolutionary relationships among ammonia-oxidizing and nitrite-oxidizing bacteria. *J. Bacteriol.* **176**, 6623–6630 (1994).
26. Kuai, L., Nair, A. A. & Polz, M. F. Rapid and simple method for the most-probable-number estimation of arsenic-reducing bacteria. *Appl. Environ. Microbiol.* **67**, 3168–3173 (2001).
27. Wenzel, W. W. *et al.* Arsenic fractionation in soils using an improved sequential extraction procedure. *Anal. Chim. Acta* **436**, 309–323 (2001).
28. McArthur, J. M., Ravenscroft, P., Safiulla, S. & Thirlwall, M. F. Arsenic in groundwater: testing pollution mechanisms for sedimentary aquifers in Bangladesh. *Wat. Resour. Res.* **37**, 109–117 (2001).
29. van der Peer, Y. & de Wachter, R. Treecon for windows — a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci.* **10**, 569–570 (2001).
30. Saitou, N. & Nei, M. The neighbor-joining method; a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by EPSRC, the Bangladesh Ministry of Science & Technology (Bangabandhu Fellowship to F.S.I.), The Royal Society, University of Manchester, ORS, GV Instruments and NERC. H. Rowland is thanked for XRD analysis. R. Bilsborrow and F. Mosselmann provided invaluable support in the acquisition of XAS data, which was supported by beamtime awards at Daresbury SRS by CCLRC. Fieldwork by D.C. was supported by KTH, IFCPAR and the University of Kalyani.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.L. (jon.lloyd@man.ac.uk).

Why large-scale climate indices seem to predict ecological processes better than local weather

T. B. Hallett¹*, T. Coulson², J. G. Pilkington³, T. H. Clutton-Brock¹, J. M. Pemberton³ & B. T. Grenfell¹

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

²Department of Biological Sciences, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK

³Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

* Present address: Department of Infectious Disease Epidemiology, Imperial College of Science Technology and Medicine, Norfolk Place, London W2 1PG, UK

Large-scale climatic indices such as the North Atlantic Oscillation¹ are associated with population dynamics², variation in demographic rates³ and values of phenotypic traits^{4,5} in many species. Paradoxically, these large-scale indices can seem to be better predictors of ecological processes than local climate^{5–8}. Using detailed data from a population of Soay sheep^{9,10}, we show that high rainfall, high winds or low temperatures at any time during a 3-month period can cause mortality either immediately or lagged by a few days. Most measures of local climate used by ecologists fail to capture such complex associations between weather and ecological process, and this may help to explain why large-scale, seasonal indices of climate spanning several months can outperform local climatic factors. Furthermore, we show why an understanding of the mechanism by which climate influences population ecology is important. Through simulation we demonstrate that the timing of bad weather within a period of mortality can have an important modifying influence on intra-specific competition for food, revealing an interaction between climate and density dependence¹¹ that the use of large-scale climatic indices or inappropriate local weather variables might obscure.

The impact of climatic variation on ecological processes has been the focus of discussion in ecology for nearly a century¹². After a period when ecologists believed that complex dynamics were determined primarily by density-dependent intrinsic processes¹³, recent work has shown that climatic variation can have an important role—either directly or through its interaction with density^{14,15}. It has emerged that large-scale seasonal indices of climate, such as the North Atlantic Oscillation (NAO, defined as fluctuations in sea-level air pressure between the Atlantic sub-polar low-pressure zone centred around Iceland and the sub-tropic high-pressure zone centred around the Azores¹), are remarkably good predictors of ecological variation—often better than local weather variables^{4–6,16–18} (but see refs 19–22 for examples where local weather predicts ecological processes well). The strong performance of large-scale climatic indices in predicting ecological processes is often difficult to reconcile with the proximal physiological processes that underpin them¹⁰. For example, in the food-limited population of Soay sheep (*Ovis aries*) considered here, there are two mechanisms by which winter weather influences mortality rates: first, by generating energetic costs on animals in poor condition, and second, by moderating vegetation productivity and available grazing for sheep during winter^{14,23,24}. In this system one may therefore expect short-term local climatic averages to predict mortality rates better than large-scale indices estimated over multiple months, yet some studies^{3,6,25} found that the NAO predicted mortality rates better than indexes of local monthly weather and explained approximately 20–30% of the variation in mortality. These results have presumably

arisen because the local climate measures used—typically average monthly temperature or rainfall—do not adequately capture the climatic conditions that influence sheep demographic rates^{7,8}. Identifying how climatic variables influence demographic rates in wild populations is, however, critical if the relative roles of density and climate are to be identified accurately. The challenge facing ecologists is to identify the appropriate climatic variables to use^{7,8}.

The Soay sheep population living in the Village Bay catchment of Hirta in the St Kilda archipelago has been the focus of a detailed, long-term, individual-based study since 1985 (ref. 10). During winter, daily mortality searches of the study area are conducted providing unusually detailed death-date data, accurate to a day for 83.3% of animals that are known to have died, for individuals born into the 1986–2002 cohorts. Daily weather data were collected by the meteorological office at Stornoway, an island approximately 100 km from St Kilda. From these data, the weekly total rainfall, and weekly averages of the maximum gust speed and minimum air temperature, were calculated. Comparison with local weather measurements on Hirta during the winters of 2000–02 indicates that the Stornoway data capture local conditions closely (temperature, $r = 0.92$; gust speed, $r = 0.82$; rainfall, $r = 0.78$). These demographic and climate data provide us with an opportunity to examine the link between climate and mortality in unprecedented detail.

Our aim is to identify climatic conditions associated with death in this system. We do this by restricting our analyses to mortality occurring in the 8 yr when the death date of more than 100 individuals is known, thus providing us with sufficiently large sample sizes to allow robust statistical analysis. Consequently we

concentrate on mortality in 1989 ($N = 227$), 1992 ($N = 213$), 1994 ($N = 125$), 1995 ($N = 160$), 1997 ($N = 180$), 1998 ($N = 110$), 1999 ($N = 393$) and 2002 ($N = 344$) (Fig. 1a). Although more than 150 animals died in each of these years (including those with less precise death dates) the overall mortality rate this represents ranges from 28% to 68% (Fig. 1a).

We divide the mortality records by sex and age (classified as lambs, yearlings and adults). The association between local weather and mortality was investigated by regressing weekly death counts against weekly weather variables (see Methods). To account for the possible delay in the effect of weather on mortality, we also regressed weekly death counts against weekly weather variables from one and two weeks previously. Detailed data on weights of animals in summer and on gut parasite load (approximated from faecal egg counts) exist for many individuals within this population²⁶ but inclusion of these terms in models did not alter our general conclusions: heavier individuals die later than lighter ones²⁷ and parasite-laden individuals die earlier than those that are relatively free of parasites²⁸, but weather in the weeks before death is still associated with mortality. We do not report these analyses in detail below as they diminish the sample size available to us within each year (as not all animals are caught) and the number of years available for comparison.

The timing of the pulses of mortality varies greatly between years (Fig. 1b). In each year most mortality occurs within the first 18 weeks (January to early May) but within this period the amount of death in each month varies considerably: the percentage of the total number of sheep dying in February, March and April ranges between 6–24%, 35–66% and 9–50%, respectively. The variation

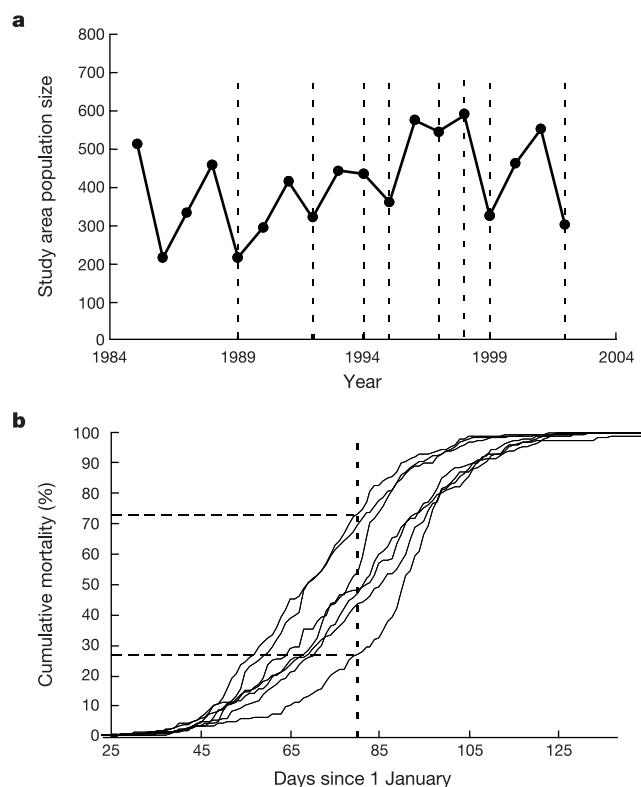


Figure 1 Dynamics of Soay sheep. **a**, Variation in study area population size between 1985 and 2002. Dotted vertical lines identify the years chosen for analysis. There were too few animals with exact death dates known to allow analysis of the 1986 population crash. **b**, Variation in the timing of mortality in different years. Each line represents the cumulative percentage mortality per day since 1 January. The vertical dotted line highlights differences in the cumulative proportion dead by 21 March, whereas the horizontal lines allow ease of reading (upper horizontal line is at 73% (1999), lower horizontal line at 27% (1994)).

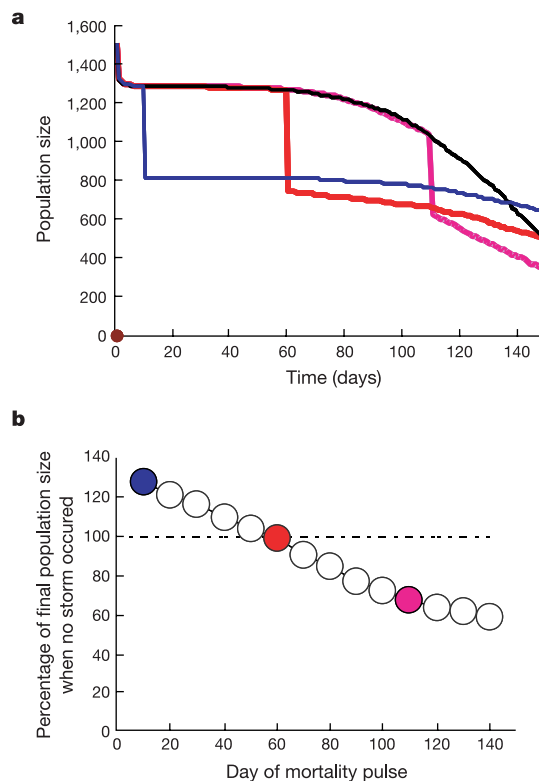


Figure 2 Results from the simulation model. **a**, Effects on population size of altering the timing of the pulse of mortality. The black line indicates the population size in the absence of a storm. The earlier the pulse is in the year (blue line, pulse at day 10; red line, pulse at day 60; purple line, pulse at day 110) the more animals survive the winter. **b**, Simulated final population size after mortality pulses as a percentage of the final population size when no pulse occurs. A pulse of mortality early in winter can generate a larger final population size than a simulation without such a pulse (dashed line). Coloured points identify simulations displayed in **a**.

in the timing of mortality helps to explain why local climatic variables calculated over a fixed 1-month time period that have previously been used²⁸ fail to predict inter-annual variation in overall mortality rates as well as the NAO. Despite variation in the timing of mortality, the overall structure of the pulse of mortality is the same in all years: males and lambs of either sex die earlier than female yearlings and female adults. The comparative lateness of female yearling and female adult mortality explains why previous research has found that February and March rainfalls are significantly associated with mortality rates in these two demographic classes.

The association between each weather variable and the timing of death varies greatly between years (Table 1). In one year a local weather variable can explain substantial amounts of variation in the timing of death, whereas in another year it can be uncorrelated with the timing of death. For example, air temperature is negatively linked to the timing of death in 2002 for almost all groups of sheep, but is largely uncorrelated with the timing of deaths in 1989 and 1997, with which rainfall and gust speed are closely associated (Table 1).

These results provide a demonstration of how the NAO can outperform local previously used weather variables^{3,6,25} in predicting ecological processes. The structure of pulses of death in each year is consistent with death by starvation (the major cause of death of sheep on Hirta²⁸) resulting from an individual's energy deficit

caused either by the energetic cost of thermoregulation or the diminished availability of forage in bad winter weather^{23,24}. The inter-annual variation in the timing of the pulses of death indicates that such conditions can occur at any time during the period from February to April (Fig. 1b). The variation in the association between each local weather variable and the timing of death (Table 1) indicates that energetically demanding conditions can occur in any form: heavy rainfall, strong wind or cold temperatures. This complex association is compounded further by the delay between these conditions occurring and the increase in mortality they cause (Table 1). Energetically challenging conditions can therefore occur at different times and in different forms in each year, and their effects may be expressed immediately or after a delay of one or two weeks. Thus, a variable in which a time and a type of energetically challenging condition is specified (for example, March rainfall) will in some years fail to capture the event that causes death. The NAO, by contrast, is a multi-month average that holds information on rainfall, wind speed and temperature¹ and thus serves as a 'catch-all' for energetically challenging conditions of any kind at any time in winter.

Local weather will, of course, always be a more powerful determinant of the timing of sheep death as long as the mechanism by which it affects ecological processes can be captured in a variable. Different components of local weather may be associated with an

Table 1 Associations between weather variables and mortality in years when substantial mortality occurred

Variable	1989			1992			1994			1995		
	W	W - 1	W - 2	W	W - 1	W - 2	W	W - 1	W - 2	W	W - 1	W - 2
Air temperature												
Male lamb	—	24%	23%	35%	—	—	—	—	25%	21%	29%	30%
Female lamb	—	28%	—	—	—	—	—	—	—	—	—	—
Male yearling	—	—	—	41%	—	—	—	49%	—	—	—	—
Female yearling	—	—	—	—	—	43%	51%	—	52%	—	—	—
Male adult	—	—	—	43%	40%	—	—	—	—	—	—	26%
Female adult	—	—	—	—	—	31%	—	—	—	—	—	—
Rainfall												
Male lamb	—	25%	35%	26%	31%	38%	—	—	—	—	—	—
Female lamb	30%	25%	—	—	—	—	—	—	—	—	—	—
Male yearling	34%	34%	—	—	—	—	—	—	—	—	47%	—
Female yearling	—	—	—	—	—	—	—	—	53%	38%	—	—
Male adult	—	37%	—	68%	22%	—	—	—	—	—	—	—
Female adult	—	20%	27%	—	—	—	—	—	—	—	—	—
Gust speed												
Male lamb	—	22%	—	—	27%	—	—	24%	—	—	—	—
Female lamb	—	32%	—	—	—	—	—	—	—	—	—	—
Male yearling	—	—	26%	—	—	—	—	—	—	—	65%	28%
Female yearling	—	—	—	—	—	—	55%	72%	44%	—	—	—
Male adult	—	—	38%	—	23%	—	—	—	—	25%	25%	41%
Variable												
1997												
1998												
1999												
2002												
Air temperature												
Male lamb	—	—	—	—	—	—	35%	30%	26%	67%	44%	47%
Female lamb	—	—	—	—	—	—	—	—	26%	49%	65%	49%
Male yearling	—	—	—	—	—	—	53%	23%	73%	—	26%	63%
Female yearling	—	—	—	—	No data	—	—	—	—	—	—	—
Male adult	22%	—	50%	28%	—	28%	—	—	58%	28%	31%	69%
Female adult	—	—	—	—	—	—	—	—	—	—	—	—
Rainfall												
Male lamb	—	—	—	—	—	32%	—	—	21%	56%	—	—
Female lamb	—	—	—	—	—	—	—	23%	—	—	—	—
Male yearling	—	—	—	—	—	—	—	—	29%	—	—	36%
Female yearling	—	—	—	—	No data	—	—	—	—	22%	—	—
Male adult	70%	—	—	—	—	—	—	—	—	27%	—	34%
Female adult	—	—	—	—	—	—	36%	—	—	—	23%	—
Gust speed												
Male lamb	—	—	23%	—	22%	—	28%	—	—	25%	—	—
Female lamb	—	—	—	—	—	—	—	—	—	—	—	—
Male yearling	—	22%	—	—	—	—	39%	41%	—	—	—	—
Female yearling	59%	—	—	—	No data	—	—	—	—	—	—	—
Male adult	—	59%	—	—	—	—	—	24%	—	36%	—	—
Female adult	—	31%	—	—	—	—	—	—	—	—	—	—

Associations with weather during the week of mortality (w) and one (w - 1) and two weeks (w - 2) before mortality are reported. Values represent the percentage of variation explained in the timing of mortality for each age group. Terms that explain more than 20% of the variation are reported; all values produce significant *P*-values following Bonferroni correction. Values in years when mortality rates in most age and sex classes were explained well by one type of weather are in bold.

ecological process in multiple ways, making the choice of a local weather variable non-trivial. Furthermore, different local ecological processes are likely to be associated with a range of local weather variables. The challenge for ecologists is to identify sensible statistics to describe the multivariate and spatially heterogeneous time series that constitute local weather. This challenge is only likely to be met if the mechanism by which climate influences ecology is known. The paradoxical performance of large-scale weather variables such as the NAO must simply reflect the low accuracy with which the biological consequences of local weather are captured with the local weather variables typically used by ecologists. In fact, for the Soay sheep system, the NAO also codes the severity of winter weather inefficiently: both high and low NAO values can reflect energetically challenging conditions¹. We expect, therefore, that detailed local weather data could be compiled into a new seasonal index that would outperform the NAO in predicting winter mortality rates.

In the Soay sheep system there are various factors that might influence mortality: the timing of a spell of bad weather, as well as previous weather in the winter. If a storm in one week killed all animals that were in poor condition an even stronger storm the following week would have little or no impact on mortality. Consequently any local measure of climate should not only incorporate multiple weather types (temperature, precipitation, wind speed) over a long period, but should also incorporate the effects of any harsh weather earlier in the season in either removing weak animals from the population or in weakening animals in better condition. Until the influence of climate can be understood in sufficient detail, the relative influence of climate and density dependence on ecological processes and the consequences of global climate change cannot adequately be appreciated. Development of a local, integrative climatic index is beyond the scope of this paper. However, we do theoretically explore how the timing of a bout of mortality could markedly modify the number of individuals that die in a year in a resource-limited population, similar to the Soay sheep system.

We constructed a simulation model of an overwintering population with a limited amount of food (see Methods). Density-dependent mortality is assumed to result from the daily intraspecific competition for food, and the chance of any individual dying is related to the amount of food it has recently consumed. Density-independent mortality is included in the form of a pulse of death (as would result from a spell of bad weather) whereby on one day each individual is exposed to an additional chance of dying. These minimal assumptions allow the results to be generalized; a consequence of this is that the model will not necessarily be a perfect caricature of any specific system, as factors such as demographic structure and day-to-day variation in energetic costs are not included in the simulation.

The later the pulse of mortality occurs, the lower the final population size (Fig. 2). This extends to the counter-intuitive result that a population that experiences an early death pulse can actually be greater in size at the end of winter than a population that does not experience any pulse (Fig. 2). The pulse of mortality, in reducing the population size, increases the chance of the remaining individuals surviving to the end of winter. This effect is enhanced when the pulse occurs earlier rather than later, as there is more food available for survivors—the victims of the death pulse having stopped eating sooner (Fig. 2). The model demonstrates that a density-dependent process can modify the strength of a density-independent process. We suggest that such time-dependent interactions may limit the power of the NAO and other crude climatic summaries in predicting ecological processes, and that a complete analysis of climatic effects in ecology cannot be limited to such data.

The ability of the NAO to outperform proxies of local climatic conditions in explaining variation in ecological processes has repeatedly been demonstrated^{4–6,16–18} and seemed paradoxical given the mechanisms by which climatic variation imposes ener-

getic stresses and influences resource availability in the wild. Our results verify that one explanation of this paradox^{7,8} may be the complicated and temporally variable associations between local climate and ecological process, which monthly climatic averages fail to capture but which the NAO can incompletely reflect. Although very few studies of free-living populations have detailed data on the timing of mortality, there is no reason to expect that these complex patterns are specific to the Soay sheep. Our results therefore provide biological justification for the inclusion of large-scale climate indices in analyses of ecological processes rather than crude summaries of local climatic conditions. However, we also stress that a detailed understanding of climate–ecology interactions can only come from synthesis of detailed data on local weather and mortality and, if possible, identification of the mechanisms by which climatic effects are expressed. □

Methods

Data

Sheep are sexed and uniquely marked within hours of birth, allowing age to be known accurately. Daily searches of the study area are conducted during winter when mortality occurs. Weather data are collected from the meteorological office at Stornoway and are available at <http://badc.nerc.ac.uk/home/>. We used data on daily precipitation (in millimetres to the nearest 0.2 mm), maximum gust speed (m s^{-1}) 2.25 m above ground and minimum ambient air temperature at 1.5 m ($^{\circ}\text{C}$).

Statistical methods

For each year analysed the week-of-death of every individual was calculated (on the basis that week 1 included death dates from 1 January to 7 January inclusive, and so on). A frequency distribution of deaths over the first 18 weeks was then calculated for male lambs (<1 yr of age), female lambs, male yearlings (>1 and <2 yr of age), female yearlings, male adults and female adults. Within-year sample sizes were not sufficient to separate male and female adults into the prime-aged and older adults age classes that have been used previously for examination of survival rates between years. Weekly climate variables were regressed against the number of deaths per week, the number of deaths the following week and the number of deaths 2 weeks later. This approach does not allow an explicit description of the 'killing power' of each weather element but does provide an estimate of the weather element with which the timing of death of those that succumb was most associated.

Simulation model

The simulation model follows the fate of N individuals through T winter days with a finite amount of food, F . Before the first day the population is divided equally among five energetic states (1–5) and the energetic class of the i th individual is denoted c_i . Population size at time t is denoted N_t . On each day the probability of each individual finding food is evaluated from a function $p(F_t, N_t)$, where F_t is the food currently available. In Fig. 2, $p(F_t, N_t) = \{1 - \exp(-aF_t)\} / \{1 + \exp(-b(N_t - N_0))\}$, where a and b are model coefficients; all functions we tried that increase monotonically over positive values of F_t and decrease monotonically over positive values of N_t produced similar results. If an individual finds food its energetic state (c_i) is increased by one; if it fails to find food its energetic state is decreased by one. If an animal in the lowest energetic state ($c_i = 1$) fails to find food it dies. The total amount of food available to the population is updated every day. The initial amount of food available is insufficient to feed each member of the initial population on every day of the simulation (that is, $F < NT$). On one day all individuals are exposed to an additional chance of dying that depends on their state on that day, $p(c_i)$. In Fig. 2, $p(c_i) = p_{\max} \exp[\beta(c_i - 1)]$, where $p_{\max}$ ($0 < p_{\max} \leq 1$) is the chance of an animal of the lowest energetic state ($c_i = 1$) dying.

Our key assumptions are that the chance of obtaining food each day depends on the current population size and the amount of food available, and that there is insufficient food to feed each member of the initial population every day. The energetic state of each individual determines its chance of dying (through both starvation and the death pulse) and this state is modified by the amount of food that has recently been consumed. The values of the parameters discussed alter the strength but not the qualitative nature of our results: in Fig. 2 $N_0 = 1,500$, $T = 150$, $F_0 = 2.10^5$, $a = 10^{-5}$, $b = 0.2$, $\beta = 0.06$ and $p_{\max} = 0.5$.

Received 12 March; accepted 24 May 2004; doi:10.1038/nature02708.

- Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
- Stenseth, N. C. *et al.* Ecological effects of climate fluctuations. *Science* **297**, 1292–1296 (2002).
- Coulson, T. *et al.* Age, sex, density, winter weather, and population crashes in Soay sheep. *Science* **292**, 1528–1531 (2001).
- Forchhammer, M. C., Clutton-Brock, T. H., Lindstrom, J. & Albon, S. D. Climate and population density induce long-term cohort variation in a northern ungulate. *J. Anim. Ecol.* **70**, 721–729 (2001).
- Post, E. & Stenseth, N. C. Climatic variability, plant phenology, and northern ungulates. *Ecology* **80**, 1322–1339 (1999).
- Catchpole, E. A., Morgan, B. J. T., Coulson, T. N., Freeman, S. N. & Albon, S. D. Factors influencing Soay sheep survival. *Appl. Stat.* **49**, 453–472 (2000).
- Mysterud, A., Stenseth, N. C., Yoccoz, N. G., Ottersen, G. & Langvatn, R. in *The North Atlantic*

- Oscillations. *Climatic Significance and Environmental Impact* (eds Hurrell, J. W., Belgrano, A., Ottersen, G. & Kushnir, Y.) 235–262 (American Geophysical Union, Washington DC, 2003).
8. Stenseth, N. C. *et al.* Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation. El Niño Southern Oscillation and beyond. *Proc. R. Soc. B* **270**, 2087–2096 (2003).
 9. Jewell, P. A., Milner, C. & Morton Boyd, J. *Island Survivors: the Ecology of the Soay Sheep of St. Kilda* (Academic, London, 1974).
 10. Clutton-Brock, T. H. & Coulson, T. N. Ungulate population dynamics: the devil is in the detail. *Phil. Trans. R. Soc.* **357**, 1299–1306 (2002).
 11. Millon, M. *Fluctuations and Order. The New Synthesis* (Springer, New York, 1995).
 12. Elton, C. S. Periodic fluctuations in the number of animals: Their cause and effects. *Br. J. Exp. Biol.* **2**, 119–163 (1924).
 13. Nicholson, A. The balance of animal populations. *J. Anim. Ecol.* **2**, 132–178 (1933).
 14. Grenfell, B. T. *et al.* Noise and determinism in synchronised sheep dynamics. *Nature* **394**, 674–677 (1998).
 15. Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton Univ. Press, Princeton, 2003).
 16. Post, E. & Stenseth, N. C. Large-scale climatic fluctuation and population dynamics of moose and white-tailed deer. *J. Anim. Ecol.* **67**, 537–543 (1998).
 17. Post, E. & Forchhammer, M. C. Synchronisation of animal population dynamics by large-scale climate. *Nature* **420**, 168–171 (2002).
 18. Stenseth, N. C. *et al.* Dynamics of coastal cod populations: intra- and intercohort density dependence and stochastic processes. *Proc. R. Soc. Lond. B* **266**, 1645–1654 (1999).
 19. Albon, S. D., Guinness, F. E. & Clutton-Brock, T. H. The influence of climatic variation on the birth weights of red deer (*Cervus elaphus*). *J. Zool.* **200**, 295–298 (1983).
 20. Albon, S. D., Clutton-Brock, T. H. & Guinness, F. E. Early development and population dynamics in red deer. 2. Density independent effects and cohort variation. *J. Anim. Ecol.* **56**, 69–81 (1987).
 21. Gaillard, J. M. *et al.* Early survival in roe deer: causes and consequences of cohort variation in two contrasted populations. *Oecologia* **112**, 502–513 (1997).
 22. Keller, L. F. & Vannoorndijk, A. J. Effects of local environmental conditions on nestling growth in the great tit *Parus major* L. *Ardea* **82**, 349–362 (1994).
 23. Clutton-Brock, T. H. *et al.* Stability and instability in ungulate populations: An empirical analysis. *Am. Nat.* **149**, 195–219 (1997).
 24. Illius, A. W. & Gordon, I. J. in *38th Symposium of the BES. Herbivores. Between Plants and Predators* (eds Oliff, H. & Drent, R. H.) (Blackwell Science, Oxford, 2002).
 25. Milner, J. M., Elston, D. A. & Albon, S. D. Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *J. Anim. Ecol.* **68**, 1235–1247 (1999).
 26. Clutton-Brock, T. H., Price, O. F., Albon, S. D. & Jewell, P. A. Early development and population fluctuations in Soay sheep. *J. Anim. Ecol.* **61**, 396–399 (1992).
 27. Clutton-Brock, T. H. *et al.* Population fluctuations, reproductive costs and life history tactics in female Soay sheep. *J. Anim. Ecol.* **65**, 675–689 (1996).
 28. Gulland, F. M. D. The role of nematode parasites in Soay sheep (*Ovis aries* L) mortality during a population crash. *Parasitology* **105**, 493–503 (1992).

Acknowledgements For permission to work on St Kilda and logistics we thank Scottish Natural Heritage, the National Trust for Scotland and the UK Ministry of Defence. For data collection we thank A. MacColl and numerous volunteers. The long-term data collection has been funded by grants from NERC, BBSRC and the Wellcome Trust (UK) to T. Clutton-Brock, B. Grenfell, J. Pemberton and M. Crawley. We also thank I. Stevenson for installing and maintaining weather stations on Hirta, and to I. Stevenson and M. Crawley for comments on an earlier draft of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.C. (t.coulson@imperial.ac.uk).

Microbialite resurgence after the Late Ordovician extinction

Peter M. Sheehan¹ & Mark T. Harris²

¹Department of Geology, Milwaukee Public Museum, 800 West Wells Street, Milwaukee, Wisconsin 53233, USA

²Department of Geological Sciences, University of Wisconsin Milwaukee, PO Box 413, Milwaukee, Wisconsin 53201, USA

Microbialites, including biogenic stromatolites, thrombolites and dendrolites, were formed by various microbial mats that trapped and bound sediments or formed the locus of mineral precipitation¹. Microbialites were common and diverse during the Proterozoic^{2–4}, but declined in abundance and morphological diversity when multicellular life diversified during the Cambrian

Radiation. A second decline occurred during the Ordovician Radiation of marine animals, and from then until the present microbialites have been confined largely to high-stress environments where multicellular organisms are rare. The microbialite declines in the Phanerozoic are attributed to disruption of the mats by animals^{2,5,6}. A resurgence of stromatolite abundance and size during reduced animal diversity after the Permian extinction⁷ has been documented anecdotally. Here we show, with statistical support, that a microbialite resurgence also occurred after the Late Ordovician extinction event in western North America. The resurgences were associated with loss of mat-inhibiting animals, providing insights into shallow-water community structures after extinction events.

We document three aspects of microbialite resurgence that followed the Late Ordovician extinction event (LOEE) using centimetre-scale sedimentological logs of carbonate rocks deposited on the shelf margin of the western USA. First, microbialites increased in size and abundance. Second, morphological diversity increased as thrombolites and columnar stromatolites appeared in abundance for the first time since the Ordovician Radiation; these earliest Silurian microbialites resemble those of the Cambrian and early Ordovician, before the macrofauna expanded during the Ordovician Radiation. Third, the microbialite resurgence, which lasted roughly 5 million years (Myr), corresponds with widespread, low-diversity megafaunal communities of the post-extinction recovery interval.

One of the most extensive tropical carbonate platforms known in the world during the Late Ordovician and Early Silurian is preserved in the eastern Great Basin. The westward-deepening carbonate platform (Fig. 1) has five upper Ordovician ramp sequences (O1–O5), and three lower-middle Silurian ramp sequences (S1–S3), followed by three rimmed-shelf sequences (S4–S6)^{8,9}. The faunas and ecological patterns have been documented^{10,11} and tied to a sequence stratigraphic framework.

Sequence O5 was deposited during an interval of glacio-eustatic sea-level draw-down associated with the LOEE, one of the five great Phanerozoic extinction events¹². The LOEE was caused by the brief (~0.5 Myr) Hirnantian glaciation that produced climatic changes

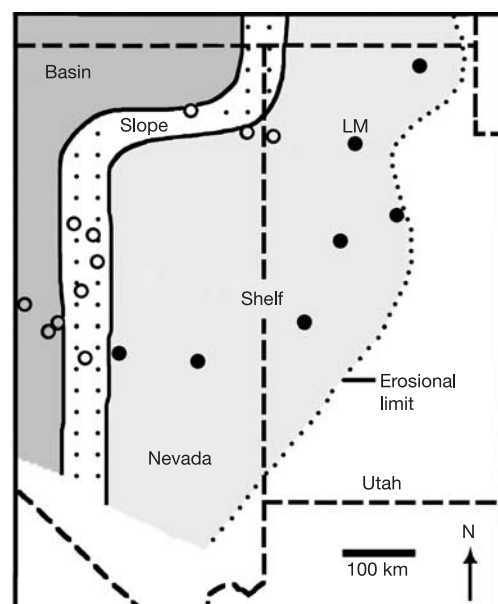


Figure 1 Palaeogeographic map of lowest Silurian (sequence S1). Circles indicate measured sections; filled circles indicate the occurrence of at least one bed with microbial domes. LM, Lakeside Mountains locality.

and loss of habitat owing to glacio-eustatic sea-level decline^{10,13}. The LOEE caused nearly total extinction of the fauna in the study region¹¹ during the deposition of sequence O5. Because the shelf was drained, sequence O5 is present only in shelf-margin, slope and basin settings. Rapid postglacial deepening marks the beginning of sequence S1, and the post-extinction faunal recovery occurred during sequences S1 and S2 (refs 8–11).

Inner-shelf facies were deposited in restricted tidal flat systems that prograded across the shelf below each sequence boundary. Stratiform microbial laminites are common and widespread indicators of intertidal or supratidal subenvironments in the tidal flats, and some domal stromatolites with heights of up to 20 cm occur in the subtidal parts of the tidal flat system (laminite facies). Domal stromatolites attain much greater heights (~60 cm; Fig. 2a) in the first two Silurian sequences than in other sequences. Furthermore, columnar stromatolites (up to ~70 cm; Fig. 2b) and thrombolites (up to ~2 m high; Fig. 2c, d) occur only in sequences S1 and S2. Many microbialites in this interval have dark rinds (Fig. 2), which are similar to *Renalcis* rinds on early Palaeozoic microbialite domes, but pervasive dolomitization inhibits petrographic study.

Detailed centimetre-scale sedimentological logs made through seven shelf sections (Fig. 1) that average ~500 m in thickness allow a statistical assessment of microbialite distributions. Sequence O5 was not included in the analysis because it was deposited only in platform margin settings that were not conducive to microbial mat growth. Similarly, the youngest three Silurian sequences (S4–S6) were not included in the analysis because they followed a change in depositional style to a rimmed-shelf profile that had few intervals with laminite facies in the region studied. Because no microbialite domes were found in the excluded intervals, inclusion of these intervals would only strengthen our conclusions.

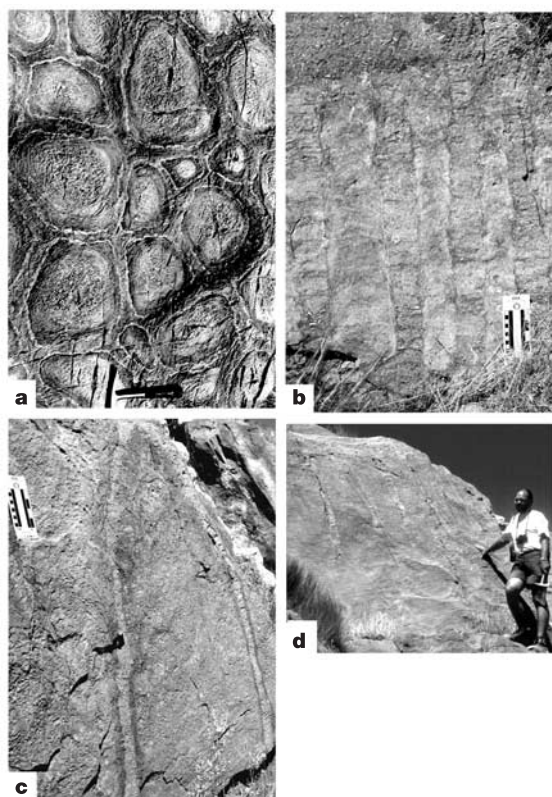


Figure 2 Microbialites from sequence S2 in the southern Lakeside Mountains, Utah. **a**, Domal stromatolites viewed from above in an erosional cross-section. **b**, Columnar stromatolites. **c**, **d**, Domal thrombolites.

Examination of more than 60 additional shelf sections complements and supports these results. Beds with domal microbialites (of any size) are significantly more abundant per 100 m of laminite facies (Table 1) in the recovery interval (sequences S1 and S2) than in the other five platform sequences on the basis of the randomization test for two independent samples¹⁴ ($n_1 = 2$, $n_2 = 5$; $P = 0.0476$). Large (>50 cm) microbialite domes are restricted to sequences S1 and S2, and bootstrapping analysis indicates that this distribution is highly non-random ($P < 0.01$; Fig. 3). In sequence S3, microbialites returned to pre-extinction patterns. Morphological diversity declined with the loss of thrombolites and columnar stromatolites. The size of domal stromatolites declined to heights of less than 20 cm, and abundance declined to pre-extinction levels.

The evolutionary significance of the resurgence lies in links between the history of microbial structures, the rise of the evolutionary faunas¹⁵ and the ecological role of mass extinctions in evolution^{10,15–17}. Microbialites dominated marine substrates during the Proterozoic but declined in abundance in the Vendian^{1–4,6,18}. The two microbialite declines in the early Phanerozoic coincided with radiations of the Cambrian and Palaeozoic evolutionary faunas^{1,3,15}. During the Early Cambrian Radiation¹⁹ microbialite abundance declined, but microbialites were still common locally during the Cambrian and Early Ordovician^{1,3,6,20}. Thrombolitic textures appeared in the early Cambrian, when animals evolved that could burrow into the mats⁶. Microbialites abundance declined again during the Ordovician Radiation of the Palaeozoic evolutionary faunas at the end of the Early Ordovician^{1,3}. Thrombolites, dendrolites and columnar (rather than domal) stromatolites nearly disappeared in level bottom settings^{1,3}. After the Ordovician Radiation, microbialites were common only in reefs²¹ or in harsh environments, where they were buffered from the effects of metazoans^{1–6,22,23}.

Microbialite declines in non-reef settings after the Cambrian and Ordovician Radiations are tied to the diversified biota's activity, such as feeding on the microbial mats, disturbing the microbial mats by movements of animals, and competition for space^{1–4}. In addition, the diversification of shelled animals may have changed oceanic chemistry sufficiently to inhibit microbialites from calcifying and stabilizing their substrate^{1,4,16,23}. The Cambrian Radiation produced an enormous increase in faunal activity¹⁹. The Ordovician Radiation gave rise to the Palaeozoic evolutionary faunas, which

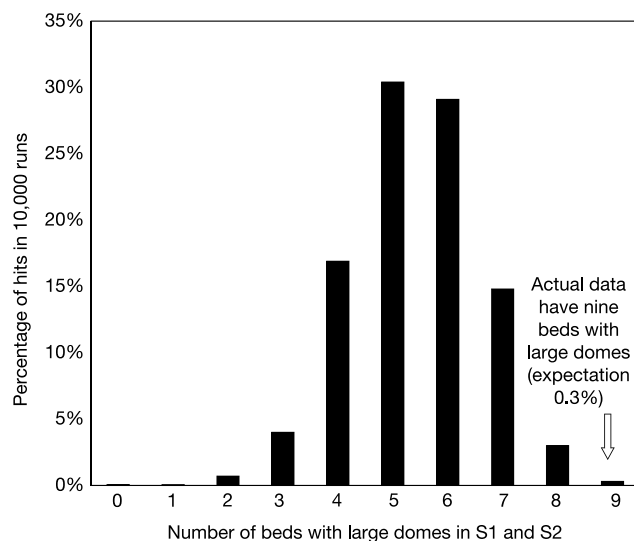


Figure 3 Bootstrapping test to determine the number of large microbial domes expected in sequences S1 and S2. The actual data comprise 9 beds with large microbial domes out of 25 beds with microbial domes of all sizes, with 15 of the beds occurring in sequences S1 and S2. The actual distribution (all 9 beds with domes >50 cm high are confined to sequences S1 and S2) occurred in only 0.3% of bootstrapping runs.

Table 1 Distribution of microbial domes relative to thickness of laminite facies* in each stratigraphic sequence

Stratigraphic sequences	Beds with microbial domes		Thickness (m)†	Laminite facies beds with domes per 100 m
	Total beds	Beds with domes >0.5 m		
S3	3	0	49	6.1
S2	11	7	77	14.3
S1	4	2	27	14.8
O4	2	0	50	4.0
O3	0	0	35	0.0
O2	4	0	118	3.4
O1	1	0	70	1.4
Total	25	9		

*The laminite facies was deposited in supratidal to subtidal parts of the tidal flat system.

†Thickness of laminite facies is the total from the seven measured sections.

were more diverse and had more complex ecological interactions (including newly evolved colonial organisms such as corals and bryozoans that were competing for substrate space), and a more mobile benthos than the Cambrian evolutionary faunas^{15,24}.

The LOEE occurred when organisms had adapted to a long greenhouse interval^{10,13}. About 30% of families and more than 50% of genera of marine invertebrates were lost¹⁵, and this extent of taxonomic extinction indicates that species diversity probably declined in excess of 80% (refs 12, 15). In addition, surviving species lost large portions of their populations. Loss of diversity was accompanied by degradation and destruction of ecological structures, including loss of whole faunal provinces^{10,25}.

During the recovery interval (S1 and S2) in the Great Basin, within-community diversity was very low and there were far fewer communities than before the extinction^{10,11}. For example, in sequences S1 and S2 only low-diversity *Virgiana* brachiopod and coral-stromatoporoid communities were common in a broad range of environments from very shallow subtidal to below storm-wave base^{10,11}. Down-ramp environments well below storm-wave base were nearly devoid of shelly fossils. This contrasts with the Late Ordovician and post-recovery Silurian, when these same physical settings hosted a wide variety of communities across a transect from shallow to slope environments¹¹. These communities were much more diverse and had more limited geographic extent, suggesting that they had more interactive species than did recovery interval communities.

Synecological conditions in the post-extinction recovery interval share similarities with the Early Ordovician: namely, low faunal diversity, wide-ranging communities and abundant microbialites²⁴. The resurgence of large domal stromatolites, thrombolites and columnar stromatolites throughout sequences S1 and S2 suggests that microbial-mat-inhibiting animals had not recovered. By sequence S3, microbialites were restricted once more by a diversified biota.

The Early Triassic stromatolite resurgence that followed the end-Permian extinction also has been attributed to a depleted biota^{7,26,27}. In addition to microbial stromatolites, the Early Triassic has massive precipitated carbonate mounds unlike anything found in this study. Abundant stromatolites have been reported in association with the Late Devonian mass extinction²⁸ and locally after the Messinian salinity crisis in the western Mediterranean^{23,29}. Thus, microbialite resurgence in association with low diversity ecosystems may be a common feature of post-extinction biotas.

Our study provides a window into the unusual ecology of post-extinction recovery intervals during which macroevolutionary change and ecological restructuring occurred. The coincidence of microbialite resurgences with intervals of faunal recovery supports the contention that extinction events reduced ecological constraints on evolution and thereby promoted rapid evolutionary changes among the survivors^{16,17}. Therefore, many evolutionary innovations took place during intervals when the ecology was very different from modern times. The current diversity crisis in the oceans³⁰ could produce a resurgence of microbialites in shallow-water environ-

ments. The appearance of microbialites might provide a local warning sign of an impending ecological crisis. □

Received 8 April; accepted 14 May 2004; doi:10.1038/nature02654.

1. Riding, R. Microbial carbonates: the geological record of calcified bacterial–algal mats and biofilms. *Sedimentology* **47**(Suppl. 1), 179–214 (2000).
2. Awramik, S. M. Precambrian columnar stromatolite diversity: reflection of metazoan appearance. *Science* **216**, 171–173 (1982).
3. Awramik, S. M. & Sprinkle, J. Proterozoic stromatolites: the first marine evolutionary biota. *Hist. Biol.* **13**, 241–253 (1999).
4. Grotzinger, J. P. & Knoll, A. H. Stromatolites in Precambrian carbonates: evolutionary mileposts or environmental dipsticks? *Ann. Rev. Earth Planet. Sci.* **27**, 313–358 (1999).
5. Garrett, P. Phanerozoic stromatolites: noncompetitive ecologic restriction by grazing and burrowing animals. *Science* **169**, 171–173 (1970).
6. Walter, M. R. & Heys, G. R. Links between the rise of Metazoa and the decline of stromatolites. *Precamb. Res.* **29**, 149–174 (1985).
7. Schubert, J. K. & Bottjer, D. J. Aftermath of the Permian–Triassic mass extinction event: paleoecology of Lower Triassic carbonates in the western USA. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **116**, 1–39 (1995).
8. Harris, M. T. & Sheehan, P. M. in *Paleozoic Sequence Stratigraphy: Views from the North American Craton* Special Paper 306 (eds Witzke, B. J., Ludvigson, G. A. & Day, J. E.) 161–176 (Geol. Soc. Am., Boulder, CO, 1996).
9. Harris, M. T. & Sheehan, P. M. in *Silurian Cycles: Linking Dynamic Stratigraphy with Atmospheric and Oceanic Changes* Bulletin 491 (eds Landing, E. & Johnson, M. E.) 51–61 (New York State Museum, Albany, NY, 1998).
10. Sheehan, P. M. The Late Ordovician mass extinction. *Ann. Rev. Earth Planet. Sci.* **29**, 331–364 (2001).
11. Sheehan, P. M. & Harris, M. T. in *Early Paleozoic Biochronology of the Great Basin, Western United States* Professional Paper 1579C (ed. Taylor, M. E.) 85–115 (US Geol. Surv., Washington DC, 1997).
12. Raup, D. M. & Sepkoski, J. J. Jr Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503 (1982).
13. Brenchley, P. J. *et al.* Bathymetric and isotopic evidence for a short-lived Late Ordovician glaciation in a greenhouse period. *Geology* **22**, 295–298 (1994).
14. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York, 1956).
15. Sepkoski, J. J. Jr Biodiversity: past, present, and future. *J. Paleontol.* **71**, 533–539 (1997).
16. Bottjer, D. J., Droser, M. L., Sheehan, P. M. & McGhee, G. R. Jr in *Evolutionary Paleobiology* (eds Allmon, W. D. & Bottjer, D. J.) 35–61 (Columbia Univ. Press, New York, 2001).
17. Jablonski, D. & Sepkoski, J. J. Jr Paleobiology, community ecology, and scales of ecological pattern. *Ecology* **77**, 1367–1378 (1996).
18. Grotzinger, J. P. Geochemical models for proterozoic stromatolite decline. *Am. J. Sci.* **290A**, 80–103 (1990).
19. Webby, B. D., Paris, F. & Droser, M. L. (eds) *The Ecology of the Cambrian Radiation* (Columbia Univ. Press, New York, 2001).
20. Bottjer, D. J., Schubert, J. K. & Droser, M. L. in *Biotic Recovery from Mass Extinction Events*. Special Pub 102 (ed. Hart, M. B.) 1–13 (Geol. Soc. London, London, 1996).
21. Wood, R. *Reef Evolution* 414 (Oxford Univ. Press, Oxford, 1999).
22. Browne, K. M., Golubic, S. & Lee, S.-J. in *Microbial Sediments* (eds Riding, R. E. & Awramik, S. M.) 233–249 (Springer, Berlin, 2000).
23. Pope, M. C., Grotzinger, J. P. & Schreiber, B. C. Evaporitic subtidal stromatolites produced by *in situ* precipitation: textures, facies associations, and temporal significance. *J. Sedim. Res.* **70**, 1139–1151 (2000).
24. Webby, B. D., Paris, F. & Droser, M. L. (eds) *The Great Ordovician Biodiversification Event* (Columbia Univ. Press, New York, 2004).
25. Sheehan, P. M., Coorrough, P. J. & Fastovsky, D. E. in *The Cretaceous–Tertiary Event and Other Catastrophes in Earth History*. Special Paper 307 (eds Ryder, G., Fastovsky, D. & Gartner, S.) 477–489 (Geol. Soc. Am., Boulder, CO, 1996).
26. Lehmann, D. J., Wei, J. & Enos, P. Controls on facies architecture of a large Triassic carbonate platform: the Great Bank of Guizhou, Nanpanjiang Basin, South China. *J. Sedim. Res.* **68**, 311–326 (1998).
27. Kershaw, S., Zhang, T. & Lan, G. A microbialite crust at the Permian–Triassic boundary in south China and its palaeoenvironmental significance. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **146**, 1–18 (1999).
28. Stephens, N. P. & Sumner, D. Y. Famennian microbial reef facies, Napier and Oscar Ranges, Canning Basin, western Australia. *Sedimentology* **50**, 1283–1302 (2003).
29. Martin, J. M. & Braga, J. C. Messinian events in the Sorbas Basin in southeastern Spain and their implications in the recent history of the Mediterranean. *Sedim. Geol.* **90**, 257–268 (1994).
30. Watson R. T. & Core Writing Team (eds) *3rd Assessment Report of the Inter-Governmental Panel on Climate Change. Climate Change 2001: Synthesis Report* [online] (<http://www.ipcc.ch/pub/syngem.htm>) (2001)

Acknowledgements J. Awe was an astute volunteer during four summers of fieldwork; and C. Morse provided statistical advice. The research was supported by National Science Foundation grants to P.M.S. and M.E.H. This is a contribution to IGCP Project 503.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.M.S. (sheehan@uwm.edu).

Local sleep and learning

Reto Huber¹, M. Felice Ghilardi², Marcello Massimini¹ & Giulio Tononi¹

¹Department of Psychiatry, University of Wisconsin, Madison, Wisconsin 53719, USA

²Center for Neurobiology and Behavior, Columbia College of Physicians and Surgeons, New York, New York 10032, USA

Human sleep is a global state whose functions remain unclear. During much of sleep, cortical neurons undergo slow oscillations in membrane potential, which appear in electroencephalograms as slow wave activity (SWA) of <4 Hz¹. The amount of SWA is homeostatically regulated, increasing after wakefulness and returning to baseline during sleep². It has been suggested that SWA homeostasis may reflect synaptic changes underlying a cellular need for sleep³. If this were so, inducing local synaptic changes should induce local SWA changes, and these should benefit neural function. Here we show that sleep homeostasis indeed has a local component, which can be triggered by a learning task involving specific brain regions. Furthermore, we show that the local increase in SWA after learning correlates with improved performance of the task after sleep. Thus, sleep homeostasis can be induced on a local level and can benefit performance.

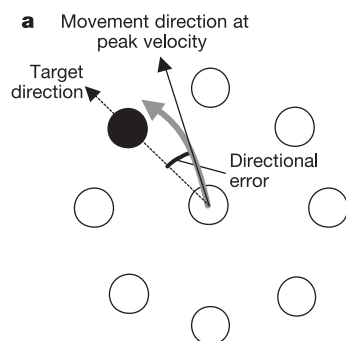
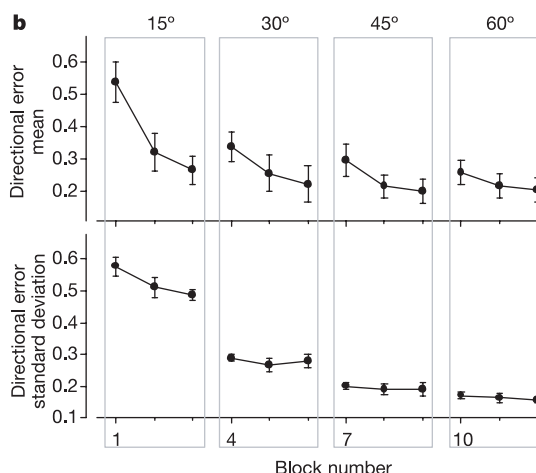


Figure 1 Rotation adaptation. **a**, The tasks. Subjects moved a handheld cursor on a digitizing tablet, executing out-and-back movements from a central starting point to one of eight targets (distance of 4.2 cm) displayed on a computer screen together with the cursor position. An opaque shield prevented subjects from seeing their arm and hand at all times. Targets were randomly highlighted at regular 1-s intervals. In the rotation adaptation task, unbeknown to the subjects, the cursor position was rotated anticlockwise relative to the hand position by a fixed angle. In a separate session, one week earlier or one week later, subjects performed the same task without any imposed rotation (no-rotation task). As shown, we computed the directional error for each movement as the angle between the line from the initial hand position to the position of the target (dotted line) and the line to

To investigate local aspects of sleep regulation, we asked subjects to perform a motor learning task just before going to sleep. In this task, subjects reach for visual targets using a handheld cursor while unconsciously adapting to systematic rotations imposed on the perceived cursor trajectory (Fig. 1a; see also ref. 4). This rotation adaptation task was chosen because it is an implicit learning paradigm; it is suitable for extended sessions; it permits accurate parameterization of both performance improvement and noise reduction; it can be contrasted with a no-rotation task that has the same kinematic requirements and is subjectively indistinguishable; and in contrast with the no-rotation task, it activates circumscribed brain regions (that is, right parietal areas 40 and 7 (ref. 4)). Over a number of movements, all our subjects adapted to the imposed rotation by progressively reducing the directional error of their trajectory as well as its variance (Fig. 1b).

Immediately after the rotation adaptation task, we recorded the sleep electroencephalogram (EEG) using a 256-channel system (Electrical Geodesics) inside a soundproofed room. As whole-night recordings with the high-density EEGs are not comfortable, the cap was removed after 2 h and subjects were then allowed to sleep undisturbed for the rest of the night (all reported satisfactory, restful sleep). As a control condition, 1 week earlier or later, subjects performed the no-rotation task, where the requirements were kinematically identical but the cursor was not rotated^{4,5}. High-density recordings after both tasks showed the usual progression of sleep stages. Values (mean $\pm$ s.e.m.) for rotation and no-rotation tasks were, respectively: sleep latency, 6.1 ± 0.8 and 6.5 ± 1.4 min; wakefulness, $14.4 \pm 4.8\%$ and $14.2 \pm 3.8\%$; rapid eye movement (REM) sleep, $2.3 \pm 0.9\%$ and $2.9 \pm 1.2\%$; non-REM sleep (stages 2–4), $65.2 \pm 6.4\%$ and $65.7 \pm 7.0\%$ (% of recording time).

Average power spectra of consecutive 4-s epochs during non-REM sleep showed that SWA was prevalent in anterior regions, in accord with previous studies^{2,6}, and that the topographic pattern was highly reproducible in both the rotation and no-rotation conditions (Fig. 2a). However, when we compared the two conditions, we found that the rotation condition produced a local increase in SWA at a cluster of six right parietal electrodes, which



the position of the hand at the peak outward velocity (solid line)⁴. The grey line represents the hand trajectory. Other movement parameters such as movement time were also computed⁵. **b**, Learning curves for the rotation adaptation task. To measure the time course of rotation adaptation, directional errors were normalized by the angle of the imposed rotation. There were four incremental steps of 15°, up to a maximum of 60° total, with short breaks in between; three blocks per step of 90 movements each (indicated at the top of the diagram). The mean directional error and the mean standard deviation for each block of 90 movements are plotted in the top and bottom panels, respectively. Points are means across subjects and bars represent standard errors. Both the mean and the standard deviation of the directional error decreased during the training blocks.

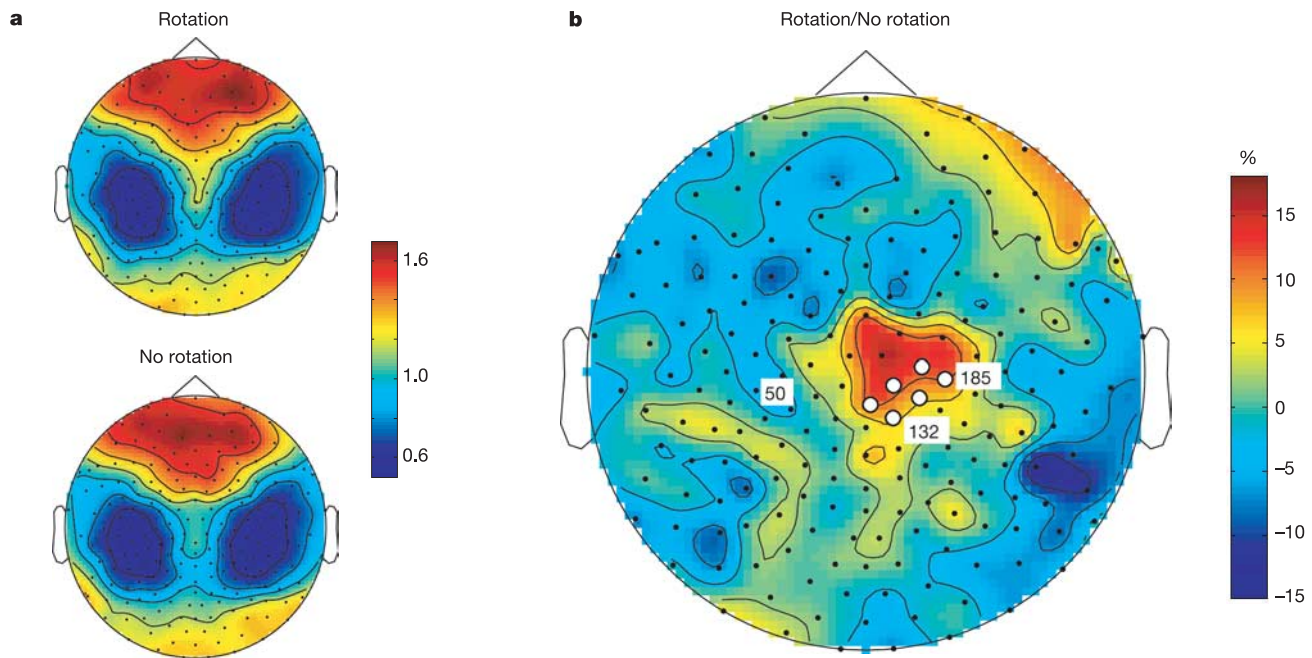


Figure 2 Local SWA homeostasis during sleep after rotation adaptation. **a**, Topographic distribution of SWA after the rotation and no-rotation tasks. Average EEG power density at 1–4 Hz (SWA, $n = 11$ subjects) for the first 30 min of non-REM sleep. EEG signals (average reference) were digitized at 500 Hz together with electromyogram and electro-oculogram, filtered (0.5–50 Hz), artefact-rejected, and sleep-staged⁶. Values (colour bar) were normalized by total power for the recording and plotted at the corresponding position

on the planar projection of the scalp surface, and interpolated (bi-harmonic spline) between electrodes (dots). **b**, Topographic distribution of the percentage change in SWA during non-REM sleep between the rotation and the no-rotation condition. White dots indicate the cluster of six electrodes showing increased SWA after rotation adaptation ($P < 0.01$, statistical nonparametric mapping, supra-threshold cluster test controlling for multiple comparisons⁷).

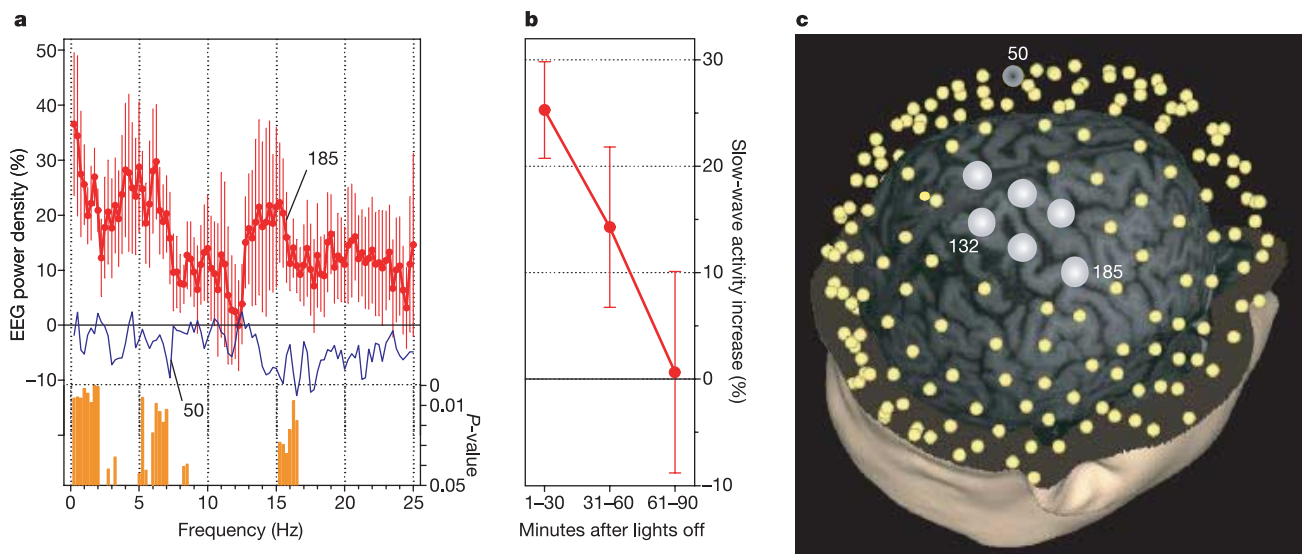


Figure 3 Frequency specificity, time course and anatomical localization of local SWA homeostasis. **a**, Frequency specificity of power changes. EEG power density spectrum for the first 30 min of non-REM sleep. Values represent the percentage change of the rotation adaptation task with respect to the no-rotation task (mean $\pm$ s.e.m. for 0.25-Hz bins, $n = 11$ subjects). Red curve, average power change across subjects for the electrode yielding the peak SWA increase for each subject in the region surrounding electrode 185; blue curve, average power change at a contralateral position, symmetric to electrode 185 (electrode 50). Bottom bars indicate frequency bins for which power in the rotation condition differed significantly from the no-rotation condition (paired t -test). **b**, Time course of SWA changes after the rotation adaptation task. The change in average EEG power in the 1–4-Hz band was calculated for three consecutive 30-min intervals during

the first non-REM sleep episode (mean $\pm$ s.e.m.). As in **a**, we selected the electrode yielding the peak SWA increase for each subject in the region surrounding electrode 185. The decline in power across the three 30-min intervals was significant ($P < 0.05$, analysis of variance). **c**, Anatomical localization of electrodes showing a significant difference in SWA during the first 30 min of non-REM sleep between the rotation and the no-rotation conditions. All 256 electrodes (yellow dots) were digitized and co-registered with the subject's magnetic resonance images. The cluster of six electrodes showing an increase in SWA is marked with white dots. For reference purposes, the electrode at the right anterior border of the cluster (185) projects onto area 40 (Talairach coordinates: $x = 50$, $y = -34$, $z = 41$), the electrode on the right posterior border of the cluster (132) projects onto area 7 (24, -68 , 48).

averaged $13.1 \pm 6.1\%$ and was statistically significant⁷ ($P < 0.01$, Fig. 2b). The peak SWA increase, whose precise location varied slightly between subjects, amounted to $25.3 \pm 4.6\%$. Thus, the rotation adaptation task elicited a response in the sleep EEG, and this response was local rather than global.

We then examined whether the local EEG response after rotation adaptation shared key features with the global homeostatic response observed in the sleep EEG after prolonged wakefulness^{2,6}. The EEG signature of sleep pressure is an increase in power predominantly in the frequency range of SWA (1–4 Hz)^{2,6} and a slight reduction in power in the sigma band (12–15 Hz). Figure 3a shows the percentage changes observed in sleep EEG power after the rotation adaptation task compared with the no-rotation condition. Consistent with a homeostatic response, there was an increase in power predominantly in the SWA frequency range, as well as a slight decrease in power in the sigma band. The increase was especially evident within the low delta band (<2 Hz) and at frequencies corresponding to the slow oscillation (<1 Hz)¹. Period amplitude analysis revealed that the increase of SWA at the site of the peak response was primarily due to an increased amplitude of slow waves ($22.1 \pm 9.7\%$ at 1–2 Hz), whereas their period was unchanged. Another feature of the global homeostatic response of SWA is its decline within and across consecutive sleep cycles². The local SWA response also revealed a decreasing trend across the first 90 min after lights off (Fig. 3b). Thus, both the spectral signature and the time course of local sleep changes resembled the global homeostatic response of sleep.

To localize the region of SWA increase, we used a positioning system (Nexstim) to digitize the 256 electrodes and co-register them with each subject's magnetic resonance images. The increase of SWA was localized to regions of the right parietal lobe encompassing Brodmann areas 40 and 7 (Fig. 3c). These regions corresponded to those highlighted by positron emission tomography (PET) subtractions between the rotation and no-rotation conditions in subjects scanned after training⁴. Both areas 40 and 7 receive converging visual and proprioceptive inputs, and are involved in processing sensory information relevant for spatial attention⁸. Area 7 is a

station in the dorsal visual pathways and processes spatial aspects of vision related to skilled actions⁸. The right hemisphere specificity in both EEG and PET experiments is in line with the right hemisphere specialization for spatial tasks⁹ and for spatial coordinate transformation⁸.

Recent work has shown that performance in certain perceptual, motor and categorization tasks improves after a night of sleep^{10–18}. These studies have demonstrated that the improvement is specifically due to sleep rather than to the mere passage of time or to circadian factors^{11–13}. We therefore asked whether performance in the rotation adaptation task would also be enhanced by sleep. As shown in Fig. 4a, when subjects were tested after a night of sleep, the directional error had decreased further, corresponding to a performance enhancement of $11.1 \pm 3.0\%$ above and beyond the level achieved at the end of awake training. By contrast, an independent group of subjects who trained in the morning and were re-tested after 8 h of wakefulness showed no such improvement (Fig. 4a). Because other movement parameters that also improved with training, such as movement time, did not differ between the two groups, the improvement in directional error was probably due to sleep and not to circadian differences. In summary, the EEG changes observed by comparing sleep after rotation adaptation with sleep after the kinematically identical no-rotation task were: (1) local, suggesting a cellular substrate; (2) characterized by an increase in SWA that behaved as expected of homeostatic changes in the sleep EEG; (3) localized to cortical regions known to be involved in learning rotation adaptation; and (4) followed the next day by an enhancement of performance.

Finally, we asked whether the post-sleep enhancement of performance and the increase of SWA in right parietal areas 40 and 7 were correlated. We found that the decrease of directional error was positively correlated with the peak increase in SWA (Fig. 4b; $r = 0.86$, $P < 0.005$). The positive correlation of post-sleep performance enhancement with the local increase in EEG power was specific to the SWA frequency range (1–4 Hz, data not shown). By contrast, there was no correlation between the improvement in rotation adaptation and SWA changes at other electrodes, or

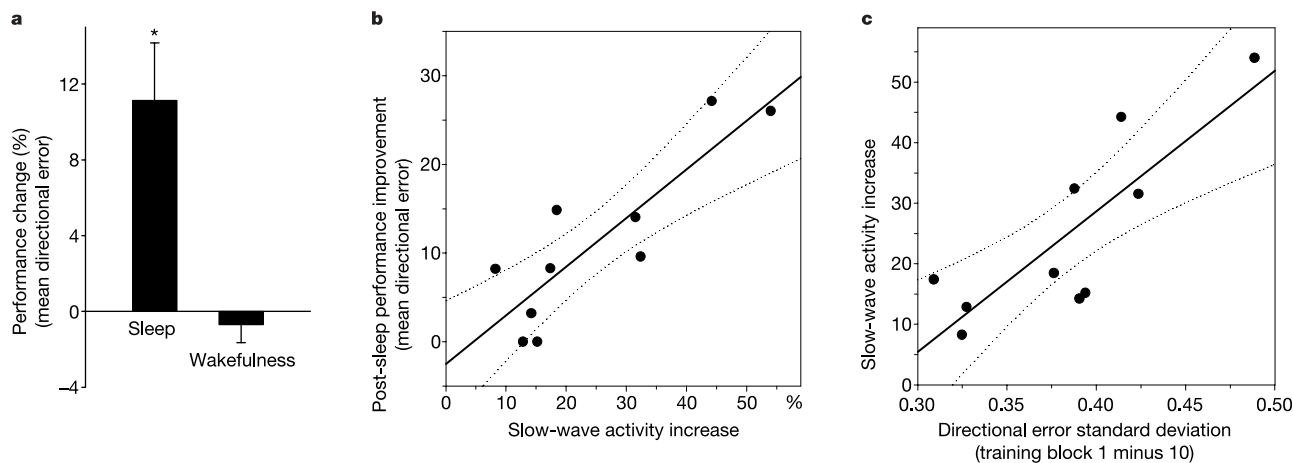


Figure 4 Enhancement of performance after sleep and its relationship to SWA.

a, Post-sleep performance improvement. Two groups of subjects (sleep group and wakefulness group) underwent rotation adaptation as described in Fig. 1. The extent of rotation adaptation was tested 10 min after the end of training using an imposed rotation of 60° . Subjects in the sleep group ($n = 11$, mean age 25.8 ± 1.8 yr) trained in the evening and were re-tested after 7–8 h of sleep (one subject did not complete the re-test and was dropped from the analysis). Subjects in the wakefulness group ($n = 10$, mean age 28.2 ± 2.9 yr) trained in the morning and were re-tested after being engaged in their normal waking activities for 8 h. A repeated measure analysis of variance and post-hoc tests showed that the two groups had similar performance when tested immediately after training; however, at re-test approximately 8 h later, mean directional error was

significantly reduced in the sleep group ($P < 0.006$) but not in the wakefulness group. The two groups differed significantly in the extent of test–re-test performance change ($P < 0.002$). **b**, Positive correlation between peak SWA increase during sleep within the six electrode cluster and decrease of mean directional error after sleep ($n = 10$ subjects). There was no correlation between the decrease of the mean directional error after sleep and the amount of total sleep or individual sleep stages during the first sleep cycle. Dotted lines indicate the 95% confidence interval. **c**, Positive correlation between the change in standard deviation of directional error during rotation adaptation (block 1 minus block 10; see Fig. 1b) and the peak SWA increase in the subsequent non-REM sleep episode ($n = 10$ subjects).

between the local increase of SWA and changes in other movement parameters such as movement time. Therefore, local SWA homeostasis in right parietal areas 40 and 7 is probably related to local neural processes specific to rotation adaptation and to their post-sleep enhancement. We did not find any correlation between the increase in SWA and the mean decrease of directional error at the end of the training session. Instead, we found a high correlation between the increase in SWA and the standard deviation of directional error at the beginning minus the end of training (Fig. 4c; $r = 0.83$, $P < 0.005$). Therefore, subjects who began rotation adaptation with much more variability or 'noise' in directional error compared to the end of training not only displayed a marked local increase in SWA during sleep but also benefited most from sleep.

These findings provide compelling evidence that the electrophysiological marker of sleep homeostasis, SWA, can be selectively induced in circumscribed regions of the cerebral cortex. Thus, they make a strong case for the local regulation of sleep^{2,6,19–22} and support a role for sleep at the cellular level^{3,23,24}. Moreover, they show that local SWA induction is triggered by a learning task, suggesting that local plastic changes associated with learning may be involved, directly or indirectly. Finally, they show that local SWA homeostasis is strongly correlated with improved performance in the task after sleep. This suggests that SWA homeostasis may be related to cellular processes underlying learning rather than to metabolic fatigue or depletion. Thus, together with evidence from intracellular studies²⁵, our results support the notion that slow oscillations might help synaptic consolidation^{25–30} or produce synaptic downscaling and increase signal-to-noise ratios in relevant neural circuits³. Most importantly, these results connect two fields that had thus far remained separate—the study of sleep homeostasis and that of sleep and plasticity. □

Received 19 March; accepted 14 May 2004; doi:10.1038/nature02663.
Published online 6 June 2004.

1. Steriade, M. Corticothalamic resonance, states of vigilance and mentation. *Neuroscience* **101**, 243–276 (2000).
2. Borbély, A. A. & Achermann, P. In *Principles and Practice of Sleep Medicine* (eds Kryger, M. H., Roth, T. & Dement, W. C.) 377–390 (W. B. Saunders, Philadelphia, 2000).
3. Tononi, G. & Cirelli, C. Sleep and synaptic homeostasis: a hypothesis. *Brain Res. Bull.* **62**, 143–150 (2003).
4. Ghilardi, M. F. *et al.* Patterns of regional brain activation associated with different forms of motor learning. *Brain Res.* **871**, 127–145 (2000).
5. Ghilardi, M. F., Eidelberg, D., Silvestri, G. & Ghez, C. The differential effect of PD and normal aging on early explicit sequence learning. *Neurology* **60**, 1313–1319 (2003).
6. Finelli, L. A., Borbély, A. A. & Achermann, P. Functional topography of the human nonREM sleep electroencephalogram. *Eur. J. Neurosci.* **13**, 2282–2290 (2001).
7. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* **15**, 1–25 (2001).
8. Cohen, Y. E. & Andersen, R. A. A common reference frame for movement plans in the posterior parietal cortex. *Nature Rev. Neurosci.* **3**, 553–562 (2002).
9. Jonides, J. *et al.* Spatial working memory in humans as revealed by PET. *Nature* **363**, 623–625 (1993).
10. Mednick, S. C. *et al.* The restorative effect of naps on perceptual deterioration. *Nature Neurosci.* **5**, 677–681 (2002).
11. Stickgold, R., James, L. & Hobson, J. A. Visual discrimination learning requires sleep after training. *Nature Neurosci.* **3**, 1237–1238 (2000).
12. Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A. & Stickgold, R. Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* **35**, 205–211 (2002).
13. Fenn, K. M., Nusbaum, H. C. & Margoliash, D. Consolidation during sleep of perceptual learning of spoken language. *Nature* **425**, 614–616 (2003).
14. Maquet, P., Schwartz, S., Passingham, R. & Frith, C. Sleep-related consolidation of a visuomotor skill: brain mechanisms as assessed by functional magnetic resonance imaging. *J. Neurosci.* **23**, 1432–1440 (2003).
15. Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. & Sagi, D. Dependence on REM sleep of overnight improvement of a perceptual skill. *Science* **265**, 679–682 (1994).
16. Fischer, S., Hallschmid, M., Elsner, A. L. & Born, J. Sleep forms memory for finger skills. *Proc. Natl Acad. Sci. USA* **99**, 11987–11991 (2002).
17. Gais, S., Plihal, W., Wagner, U. & Born, J. Early sleep triggers memory for early visual discrimination skills. *Nature Neurosci.* **3**, 1335–1339 (2000).
18. Smith, C. Sleep states and memory processes. *Behav. Brain Res.* **69**, 137–145 (1995).
19. Vyazovskiy, V., Borbély, A. A. & Tobler, I. Unilateral vibrissae stimulation during waking induces interhemispheric EEG asymmetry during subsequent sleep in the rat. *J. Sleep Res.* **9**, 367–371 (2000).
20. Kattler, H., Dijk, D.-J. & Borbély, A. A. Effect of unilateral somatosensory stimulation prior to sleep on the sleep EEG in humans. *J. Sleep Res.* **3**, 159–164 (1994).
21. Huber, R., DeBoer, T. & Tobler, I. Topography of EEG dynamics after sleep deprivation in mice. *J. Neurophysiol.* **84**, 1888–1893 (2000).
22. Krueger, J. M. & Obál, F. Jr A neuronal group theory of sleep function. *J. Sleep Res.* **2**, 63–69 (1993).

23. Tononi, G. & Cirelli, C. Some considerations on sleep and neural plasticity. *Arch. Ital. Biol.* **139**, 221–241 (2001).
24. Cirelli, C., Gutierrez, C. M. & Tononi, G. Extensive and divergent effects of sleep and wakefulness on brain gene expression. *Neuron* **41**, 35–43 (2004).
25. Steriade, M. & Timofeev, I. Neuronal plasticity in thalamocortical networks during sleep and waking oscillations. *Neuron* **37**, 563–576 (2003).
26. Maquet, P. The role of sleep in learning and memory. *Science* **294**, 1048–1052 (2001).
27. Hoffman, K. L. & McNaughton, B. L. Sleep on it: cortical reorganization after-the-fact. *Trends Neurosci.* **25**, 1–2 (2001).
28. Stickgold, R., Hobson, J. A., Fosse, R. & Fosse, M. Sleep, learning, and dreams: off-line memory reprocessing. *Science* **294**, 1052–1057 (2001).
29. Benington, J. H. & Frank, M. G. Cellular and molecular connections between sleep and synaptic plasticity. *Prog. Neurobiol.* **69**, 71–101 (2003).
30. Walker, M. P. A refined model of sleep and the time course of memory formation. *Behav. Brain Sci.* (in the press).

Acknowledgements We thank C. Cirelli, F. Ferrarelli and T. Shakhnovich for their help, colleagues at WISPIC and Columbia for their comments on the manuscript, and R. Davidson and A. Alexander at the Keck Center for support with EEG and MRI facilities. This work was supported by grants from the Swiss Foundation for Fellowships in Biology and Medicine to R.H., from the NINDS to M.E.G., from the National Sleep Foundation to M.M. and from the NIMH to G.T.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.T. (gtononi@wisc.edu).

Resilient circadian oscillator revealed in individual cyanobacteria

Irina Mihalcescu¹, Weihong Hsing^{2*} & Stanislas Leibler³

¹Laboratoire de Spectrométrie Physique, Université Joseph Fourier – Grenoble I, BP87, 38402 St-Martin d'Hères Cédex, France

²Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

³Laboratory of Living Matter and Centre for Studies in Physics and Biology, The Rockefeller University, 1230 York Avenue, New York 10021, USA

* Present address: Johnson & Johnson PRD, PO Box 300, Raritan, New Jersey 08869, USA

Circadian oscillators, which provide internal daily periodicity, are found in a variety of living organisms, including mammals, insects, plants, fungi and cyanobacteria¹. Remarkably, these biochemical oscillators are resilient to external and internal modifications, such as temperature and cell division cycles. They have to be 'fluctuation (noise) resistant'² because relative fluctuations in the number of messenger RNA and protein molecules forming the intracellular oscillators are likely to be large. In multicellular organisms, the strong temporal stability of circadian clocks, despite molecular fluctuations, can easily be explained by intercellular interactions^{3–5}. Here we study circadian rhythms and their stability in unicellular cyanobacteria *Synechococcus elongatus*. Low-light-level microscopy has allowed us to measure gene expression under circadian control in single bacteria, showing that the circadian clock is indeed a property of individual cells. Our measurements show that the oscillators have a strong temporal stability with a correlation time of several months. In contrast to many circadian clocks in multicellular organisms, this stability seems to be ensured by the intracellular biochemical network, because the interactions between oscillators seem to be negligible.

Cyanobacteria *S. elongatus* sp. PCC7942, chosen for the present study, is among the simplest organisms showing circadian behaviour⁶. Recent investigations of large populations of these prokaryotic cells have shown that the expression of most of their genes is under the control of a circadian clock⁷. To measure circadian gene

expression in a single cell, we used a bacterial luciferase reporter system⁸ consisting of two neutral site chromosomal insertions: *psbAI::luxAB* and *psbAI::luxCDE*. This bioluminescence system, which does not require exogenously added aldehyde substrate, had been successfully used to monitor the circadian clock in populations of cyanobacteria. Note that the use of fluorescent reporters, such as green fluorescent protein (GFP), would be very difficult to implement in *S. elongatus* cells, mainly because of their high auto-fluorescence. In addition, the metabolism and circadian clock would be affected by the strong light needed for fluorescent protein excitation⁹, and GFP could be 'bleached' by the prolonged light exposure needed for cell growth.

However, detecting and imaging individual bacteria 2.3–6 μm long requires at least 50 times higher sensitivity than previous bacterial population investigations. This approximates to the ratio between the number of cells in the smallest population yet monitored¹⁰ and a single individual. To achieve this sensitivity, we implemented an experimental set-up based on a back-illuminated cooled CCD (charge-coupled device) detection camera with high quantum efficiency (see Methods). Because the light levels obtained from a single cyanobacterium were typically of the order of 10–20 photons per minute per cell, we used relatively long integration times (30 min) to maximize the signal-to-noise ratio without significantly affecting the circadian rhythm. Computerized control of internal light and temperature, and an entirely automated data acquisition system, allowed measurements to be taken over prolonged periods (up to two weeks) under constant conditions.

In favourable conditions, the cyanobacteria can divide as fast as three to four times during a single circadian cycle¹⁰, while keeping a precise circadian rhythm for the growing population. As the circadian rhythm in cyanobacteria is independent of the cell division cycle^{11,12}, we imposed an average division time of ~ 23 h (see Methods) by applying moderate illumination. In this way we were able to observe a slow growth of micro-colonies out of single 'progenitor' cells over long time periods. Figure 1 shows an example of such micro-colony growth during the first 5.5 days (1 day = 24 h), monitored by both phase-contrast and bioluminescence microscopy. The inoculated bacterium (marked F in Fig. 1a) grew slowly, without dividing for the first ~ 1.2 days. Its density of bioluminescence (defined as the total luminescence divided by the cell size, see Methods) clearly oscillated (Fig. 1b) with a well-defined period of 25.4 ± 0.12 h (mean $\pm$ s.d.) (Fig. 1f). After 1 day, the cell doubled its size and started to divide. Figure 1c shows the time course of the size of cell F and all its progeny, extracted from the phase-contrast photographs. The measurement of the total size of all cells in the micro-colony (Fig. 1d) showed that the cells were in the exponential growth phase, with an average doubling time of 23.04 ± 0.17 h. The circadian oscillations of the siblings of the progenitor cell are depicted in Fig. 1e. All the siblings (such as four siblings obtained after two divisions of the initial cell, marked F, F1, F2 and F3 in Fig. 1) were oscillating with a striking synchronicity. Each of the cells had a different amplitude of oscillation, but they were characterized by similar period and phase. Figure 1f shows the density of bioluminescence averaged over all progeny. The average oscillation is well described by a simple periodic function, $\langle d(t) \rangle = B + A \cos(2\pi t/T_0 + \varphi_0)$, in which: d represents density of bioluminescence; $\langle d \rangle$, the average over all progeny; t , time; T_0 , period; φ_0 , initial phase; A , amplitude; and B , the offset. The average oscillation has not only a constant period and phase, but also a constant amplitude. In the same way, we analysed a few other neighbouring micro-colonies derived from individual cells (see Supplementary Information). Similarly, each progeny oscillated synchronously, with the same period for all the cells studied, and with an average phase that was characteristic of each micro-colony.

We quantified roughly the stability of the individual oscillators and their phase dispersion through successive cell divisions. For this

we used the average $\langle d(t) \rangle$ and the variance $\sigma_d^2(t) = \langle (d_i(t) - \langle d(t) \rangle)^2 \rangle$ of the signals $d_i(t)$ coming from each individual oscillator i studied. Quantitative analysis of relatively short time sequences, which include only a few oscillations, is both difficult and arbitrary. In a simple stochastic model, the bioluminescence signal from each cell can be expressed as $y_i(t) = A_i(t)[b + \cos(\omega_0 t + \varphi_i(t))]$, where the angular frequency ω_0 and the relative offset b are constant. The amplitude $A_i(t)$ and the phase $\varphi_i(t)$ of each oscillator are two random functions: $A_i(t)$ is a stationary gaussian process, with constant average $\langle A(t) \rangle = A$ and standard deviation σ_A , and $\varphi_i(t)$ is a wiener process with a constant mean $\langle \varphi(t) \rangle = \varphi_0$ and a variance growing linearly with time $\sigma_\varphi^2 = Dt$, where D is the diffusion constant¹³. By comparing the theoretical variance for such random oscillators, $\sigma_d^2(t)$, with the experimental variance, we estimated the amplitude noise, $\eta_A = \sigma_A/\langle A \rangle = 0.25 \pm 0.01$, and the phase diffusion constant, $D = 5 \times 10^{-4} \pm 3 \times 10^{-4} (\text{h}^{-1})$ (see Supplementary Information). These estimates confirm that the observed oscillators have relatively large amplitude fluctuations, but remain strongly in synchrony with a correlation time $\tau = 2/D = 166 \pm 100$ days. Stochastic effects in gene expression fluctuations, such as molecular noise, could be at the origin of these fluctuations, and a noise level of

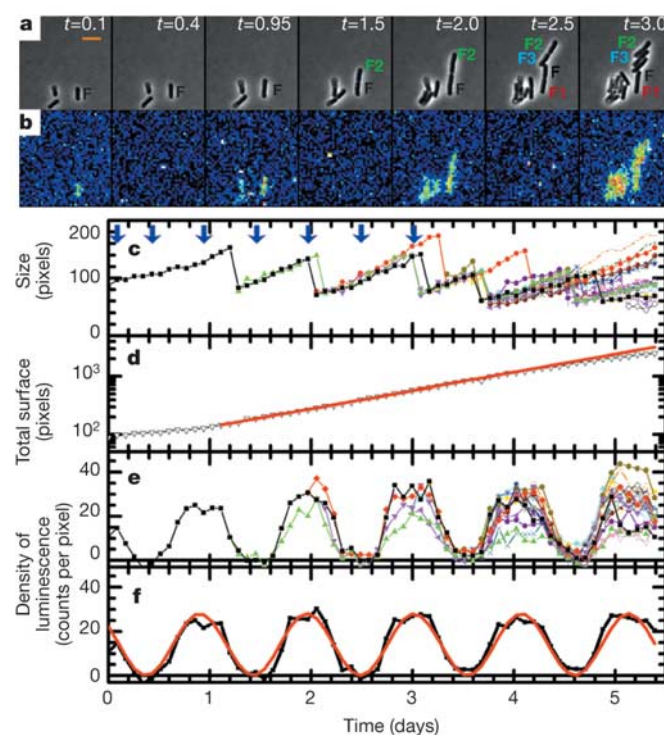


Figure 1 Circadian oscillation of bioluminescence in individual bacteria. **a**, Snapshots of phase-contrast image showing cell F and its progeny and **b**, related bioluminescence image at different times t (given in days, a 24 h period of time) from the beginning of the measurement. Pixels in the bioluminescence images were binned 3×3 (pseudo-colour, where red is high signal intensity and blue is low signal intensity). Scale bar, $5 \mu\text{m}$. **c**, The size of the cell F and all its progeny as a function of time measured from the phase-contrast images (non-binned pixels). The arrows point to the time where the snapshots in (**a**) and (**b**) were taken. **d**, The total number of pixels occupied by F and its all progeny versus time (black line) plotted in a logarithmic scale. The red line is the corresponding exponential growth fit: total size (t) = initial size $\times 2^{t/\tau}$ with $\tau = 23.04 \pm 0.17$ h. **e**, Density of bioluminescence for the same cell and all its progeny versus time. **f**, The average density of bioluminescence versus time (black line) and its fit (red line) with: $\langle d(t) \rangle = B + A \cos(2\pi t/T_0 + \varphi_0)$. The resulting period is $T_0 = 25.4 \pm 0.12$ h, the initial phase $\varphi_0 = 52 \pm 2.8^\circ$, the amplitude $A = 12.9 \pm 0.3$ counts per pixel and the offset $B = 14.8 \pm 0.3$ counts per pixel.

gene expression of $\eta = 0.25$ is not unusual¹⁴. What is surprising is the observed high level of synchronicity. It is tempting to conclude that this stability of oscillations is due to the design of the circadian genetic network in each individual cell. Another possibility, however, is that the individual oscillators become synchronized by coupling one with another. Indeed, many theoretical studies, reinforced by recent experiments¹⁵, have showed that increasing the coupling between noisy—or even chaotic—chemical oscillators results in their synchronization and the establishment of common, robust rhythms.

To test for this possibility, we followed micro-colonies growing in close proximity, but originating from individual cells that had different initial phases of circadian oscillations. Figure 2a presents an example of two such growing micro-colonies, together with the bioluminescence signal showing their out-of-phase circadian oscillations. The 'progenitor' cells, initially separated by $\sim 30 \mu\text{m}$, progressively formed two micro-colonies, which after 10–12 successive division cycles got into close contact. Figure 2b depicts the normalized density of bioluminescence of some of the descendants of the two progenitor cells (see Methods). It is quite clear that all the cells from both micro-colonies oscillate with essentially the same period, while their phase is specific to each colony. The phase of the cells in the same colony spreads slightly with time. Figure 2c represents the phases of the chosen oscillators from the two colonies (in red and black) on a circular 'phase plot'¹⁶. Each group of oscillators is also represented by a vector whose orientation angle gives the average phase and whose magnitude indicates the phase coherence (that is, its length is 1 when the oscillators are in phase, and 0 when the phase of the oscillators is uniformly randomly distributed). As we can see from this plot, the internal synchrony of each colony is close to 1 (>0.94) in the first days, and it is only

slightly diminishing (>0.85) at the end of the experiment.

We can get some insight into the origins of this slow loss of synchronicity by examining more closely the spatial evolution of the phase in growing micro-colonies. In Fig. 3, oscillations in the cells from four micro-colonies are tracked as a function of time. The temporal evolution of the phase for some of the cell lines (see Methods) has been quantified as described in Fig. 2c. We represent the average phase of the circadian oscillator in each time interval by the colour of the overlaid temporal track. It is easy to see that the difference in phases of different cell lines does not notably decrease when they are driven closer to each other (Fig. 3, arrows). Of particular interest is the situation in which the phase of oscillations in two cell lines descending from the same initial progenitor cell differs, after a few cell divisions, by a measurable amount of ~ 30 degrees (see, for example, Fig. 3, pink arrow). This phase difference then persists through numerous division cycles even though both cell lines remain physically in close contact (Fig. 3). These observations confirm that any interactions between closely packed cells seem to have negligible effect on their relative phase of oscillations, at least for phase differences large enough to be detected (see Supplementary Information for an estimate of the detection limit of the coupling strength between oscillators).

The textbook definition of circadian clocks implies the existence of self-sustained oscillations, which can be entrained to the precise 24 h period by external cues (such as light and temperature), and which are temperature-compensated. These three requirements put definite constraints on the design of biochemical networks, which can implement circadian clocks in different multicellular organisms. Here we have shown that at least for cyanobacteria, these three requirements have to be supplemented by an additional one. In these small prokaryote cells, circadian oscillators are very resilient to

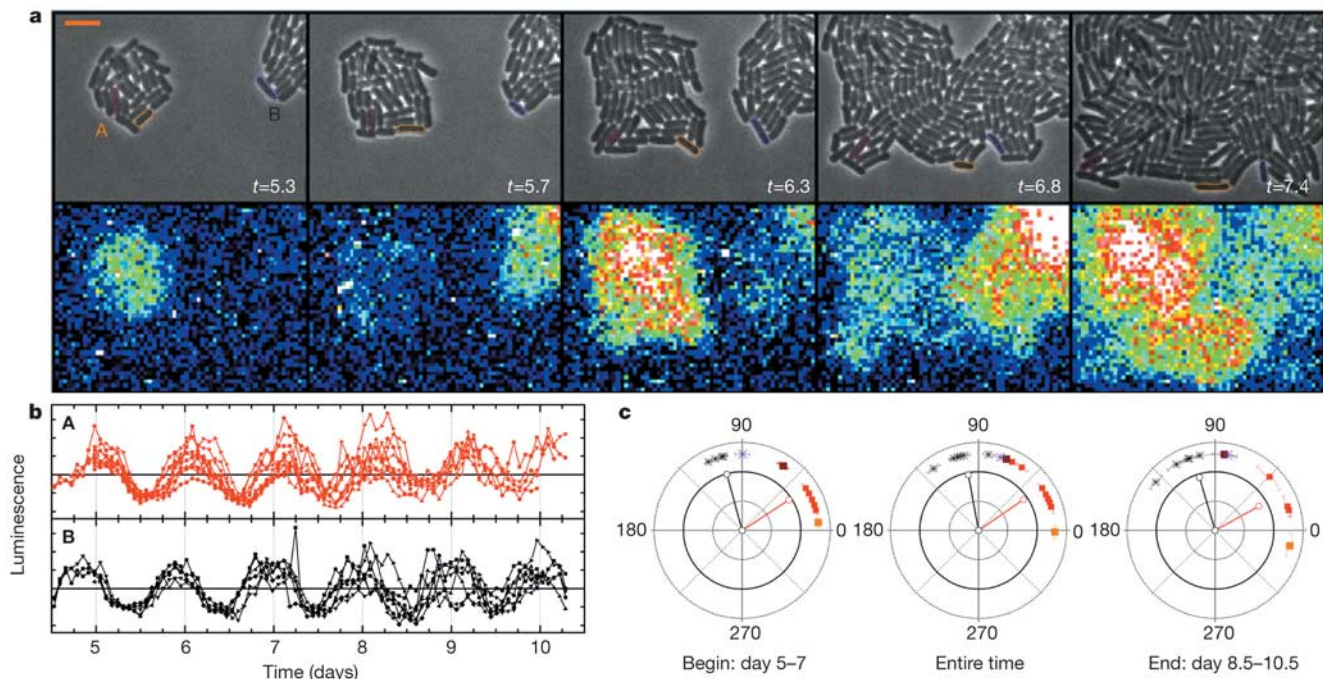


Figure 2 Growing micro-colonies of cyanobacteria, oscillating with different phases. **a**, Upper part shows the phase-contrast snapshots of colonies A and B; lower part shows the related bioluminescence images. Scale bar, $5 \mu\text{m}$. **b**, Normalized density of bioluminescence of individual cyanobacterial cells. Each colour corresponds to the progeny from one of the initial cells: red line, colony A; black line, colony B. **c**, Phase of individual oscillators as a function of their original colony and their evolution in time: red square, colony A; asterisk, colony B. An example of the exact location for three of the cells

tracked and their phase evolution is shown, marked by the corresponding coloured lines: magenta, orange and purple. The change of the phase in time was quantified by a fit over a different period of time: the first 2 days (days 5–7), the entire time (days 5–10.5) and the last 2 days of the measurement (days 8.5–10.5). The fit function is $\langle d(t) \rangle = B + A \cos(2\pi t/T_0 + \varphi)$, with $T_0 = 24.78 \text{ h}$. The line segments in each graph, with corresponding colours, represent the resulting vector $\mathbf{P}_{\text{res}} = \sum \mathbf{P}_i$, where $\mathbf{P}_i$ is the unit vector whose orientation is the measured angle of the same colony cell i .

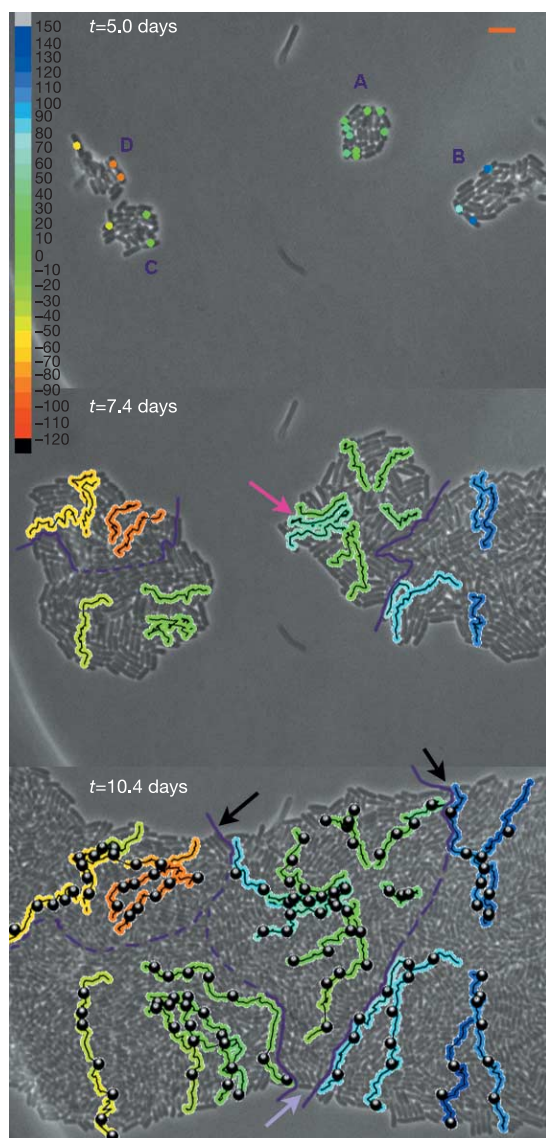


Figure 3 Temporal evolution of individual oscillators' phase is independent of close proximity. Snapshots of the four growing colonies (named A, B, C and D) in phase-contrast microscopy at three different times, t . Superposed are the tracks of the centre of gravity of each cell followed in time. The colour of each track is given by the phase of the circadian oscillation of the cell. Note that the colour of the interval 7–8.5 days of the time track represents the average phase of the given individual oscillator. The black filled dots represent a cell division event. The indigo lines show the precise (continuous line) or estimated (dashed line) boundary between the merging colonies. The arrows point to examples of spatially close cells oscillating with different phases. Scale bar, 5 μm .

perturbations associated with high intracellular noise and frequent cell divisions. The remarkable stability in time revealed in our experiments seems to be implemented at the single-cell level, and interactions between the oscillators in different cells seem to be negligible. This requirement of 'noise resistance' is a nontrivial one, and imposes strong constraints on any models of circadian biochemical networks^{17–19}.

Methods

Bioluminescence imaging

The detection system was chosen to have the highest quantum efficiency (QE) and a low electronic noise: a back-illuminated CCD system with 1,340 $\times$ 1,300 pixels, each of 20 $\mu\text{m} \times$ 20 μm , cooled to -100°C and having QE, for the spectral range of the luciferase, higher than 85% (Roper Scientific-Princeton Instruments Camera, CT 1300B Cryotiger). For imaging cells with typical sizes of 2.3–6 μm , we used a high-magnification optical

microscope with $\times 100$ objective (Zeiss Axiovert 200 with a bottom port for a direct path to the detector and an oil contact lens Plan-Neofluar 100xPh3 numerical aperture = 1.3). The data acquisition was entirely automated with custom-made software, interfaced with the camera control software, Win View/32. For automatic focusing we used phase-contrast imaging (exposure 150 ms under low-intensity bright-field light). After each focusing we kept the system in dark for 2 min to avoid the phosphorescence of the intracellular chromophores. The bioluminescence images were subject to 3×3 binning of the pixels (to diminish the read-out noise of the camera). The growth of the cells was then assured by an exposure to fluorescent lamp light for 60–90 min (depending on the experiment) with a flux of $\sim 100 \mu\text{E m}^{-2} \text{s}^{-1}$ at the surface of the sample. One field of view, preceded by an automatic focusing, was taken every 95–125 min, and the length of one experiment was 1 to 2 weeks.

Cells and growth chamber

The transformed reporter strain⁸ AMC412 was grown in BG-11 medium at 30°C with $7.5 \mu\text{l ml}^{-1}$ chloramphenicol and $5 \mu\text{g ml}^{-1}$ spectinomycin, under 12 h light/12 h dark cycles for entrainment, then under continuous light (LL) for two different periods of time: 1 week and 2 months. The cells were then diluted in fresh media at 30°C and grown again for 1 week in LL. The growth chambers for the microscope were made on 50 $\times$ 9 mm Petri dishes, with a 22 mm hole and a 25 mm No. 0 round coverslip (Erie Scientific) glued inside. Two to three microlitres of cells were spotted on the coverslip, coated with 100 μl of liquid 3% SeaPlaque low-melt agarose in media, and immediately covered and insulated with a transparent 0.4- μm -pore-size polyethylene terephthalate track-etched membrane (Falcon), and paraffin sealed. The Petri dish was then filled with 30°C media, closed, incubated at 30°C in LL for a few hours, and then installed under the microscope. In this way, the movement of the cells was highly restrained by the agarose gel, but all the cells retained homogeneous access to growth media, while a sufficiently transparent light pathway allowed phase-contrast microscopy.

Image and data analysis

Individual cells were manually identified on the phase-contrast images, from the final frame, backwards in time, to the first frame. A custom Matlab (MathWorks)-based software selected from the image the pixels corresponding to a given cell, extracted its size and position, and then in the bioluminescence image grouped the pixels 3×3 to identify the cell and to integrate its total bioluminescence. The cosmic rays, detected during the acquisition process, were easily discriminated owing to their high intensities when compared with the bacterial bioluminescence. The background value b_k , constant for each bioluminescence frame, was taken as the maximum of the intensity histogram of all individual pixels chosen in a region devoid of cells. The density of bioluminescence per cell per pixel is calculated as: $d = \sum_{i=1}^{N_T} (\text{counts})_i / N_T - b_k$, where N_T is the number of binned pixels describing the cell, and $(\text{counts})_i$ is the number of counts recorded from the pixel i . Note that both $(\text{counts})_i$ and N_T are random variables: $(\text{counts})_i$ owing to the Poisson statistics of photons and the noise of the CCD camera, and N_T because of the sampling of the continuous contour of a cell with a small number of pixels (see Supplementary Information). All fits have been done using the Levenberg–Marquardt χ^2 minimization algorithm with Origin (OriginLab) software and the uncertainty of parameters is the standard deviation obtained from the curvature of the χ^2 around its minimum.

The choice of individual cells for study was restricted to those growing essentially in a single two-dimensional layer. Note that if one cell lies on top of another cell, the bioluminescence information becomes irrelevant at the single-cell level and its track has to be aborted. Therefore, the cell lines chosen were those having the longest history in a single monolayer. This constraint was more restrictive in the last experiment. The number of cells grew ~ 50 – 100 times (that is, ~ 6 division cycles) and most of the cells soon after division were caught in multilayered colonies. At the interface between two merging colonies, the stacking up is enhanced and the cells available for tracking were typically the few that remained at the common outside border of both colonies.

Received 23 November 2003; accepted 30 March 2004; doi:10.1038/nature02533.

- Young, M. W. & Kay, S. A. Time zones: a comparative genetics of circadian clocks. *Nature Rev. Genet.* **2**, 702–715 (2001).
- Barkai, N. & Leibler, S. Circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
- Pikovsky, A., Rosenblum, M. & Kurths, J. *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge Univ. Press, Cambridge, 2001).
- Liu, C., Weaver, D. R., Strogatz, S. H. & Reppert, S. M. Cellular construction of a circadian clock: period determination in the suprachiasmatic nuclei. *Cell* **91**, 855–860 (1997).
- Yamaguchi, S. *et al.* Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science* **302**, 1408–1412 (2003).
- Kondo, T. *et al.* Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc. Natl Acad. Sci. USA* **90**, 5672–5676 (1993).
- Liu, Y. *et al.* Circadian orchestration of gene expression in cyanobacteria. *Genes Dev.* **9**, 1469–1478 (1995).
- Katayama, M., Tsinoiremas, N. F., Kondo, T. & Golden, S. S. *cpmA*, a gene involved in an output pathway of the cyanobacterial circadian system. *J. Bacteriol.* **181**, 3516–3524 (1999).
- Tsinoiremas, N. F., Schaefer, M. R. & Golden, S. S. Blue and red light reversibly control *psbA* expression in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *J. Biol. Chem.* **269**, 16143–16147 (1994).
- Kondo, T. *et al.* Circadian rhythms in rapidly dividing cyanobacteria. *Science* **275**, 224–227 (1997).
- Mori, T. & Johnson, C. H. Independence of circadian timing from cell division in cyanobacteria. *J. Bacteriol.* **183**, 2439–2444 (2001).
- Mori, T., Binder, B. & Johnson, C. H. Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl Acad. Sci. USA* **93**, 10183–10188 (1996).
- Gaspard, P. The correlation time of mesoscopic chemical clocks. *J. Chem. Phys.* **117**, 8905–8916 (2002).
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).

15. Kiss, I. Z., Zhai, Y. & Hudson, J. L. Emerging coherence in a population of chemical oscillators. *Science* **296**, 1676–1678 (2002).
16. Levine, J. D., Funes, P., Dowse, H. B. & Hall, J. C. Resetting the circadian clock by social experience in *Drosophila melanogaster*. *Science* **298**, 2010–2012 (2002).
17. Winfree, A. T. *The Geometry of Biological Time*, 2nd edn (Springer, New York, 2001).
18. Goldbeter, A. Computational approaches to cellular rhythms. *Nature* **420**, 238–245 (2002).
19. Gonze, D., Halloy, J. & Goldbeter, A. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl Acad. Sci. USA* **99**, 673–678 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Golden for the AMC412 strain and for advice, D. Peoples for advice and technical assistance, B. Houchmandzadeh, J. Paulsson, J. Vilar, C. Weitz and M. Young for discussions, and N. Questembert-Balaban, E. Kussell and M. Vallade for comments on the manuscript. This work was supported partially by Princeton University through the Lewis Thomas Fellowship (I.M.), the National Institutes of Health, the Howard Hughes Medical Institute and the Centre National de Recherche Scientifique through an ATIP and an AC 'Dynamique et réactivité des assemblages biologiques'.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.M. (imihalce@spectro.ujf-grenoble.fr).

Evolutionary changes in *cis* and *trans* gene regulation

Patricia J. Wittkopp, Belinda K. Haerum & Andrew G. Clark

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

Differences in gene expression are central to evolution. Such differences can arise from *cis*-regulatory changes that affect transcription initiation, transcription rate and/or transcript stability in an allele-specific manner, or from *trans*-regulatory changes that modify the activity or expression of factors that interact with *cis*-regulatory sequences^{1,2}. Both *cis*- and *trans*-regulatory changes contribute to divergent gene expression, but their respective contributions remain largely unknown³. Here we examine the distribution of *cis*- and *trans*-regulatory changes underlying expression differences between closely related *Drosophila* species, *D. melanogaster* and *D. simulans*, and show functional *cis*-regulatory differences by comparing the relative abundance of species-specific transcripts in F₁ hybrids^{4,5}. Differences in *trans*-regulatory activity were inferred by comparing the ratio of allelic expression in hybrids with the ratio of gene expression between species. Of 29 genes with interspecific expression differences, 28 had differences in *cis*-regulation, and these changes were sufficient to explain expression divergence for about half of the genes. *Trans*-regulatory differences affected 55% (16 of 29) of genes, and were always accompanied by *cis*-regulatory changes. These data indicate that interspecific expression differences are not caused by select *trans*-regulatory changes with widespread effects, but rather by many *cis*-acting changes spread throughout the genome.

D. melanogaster and *D. simulans* diverged about 2.5 Myr ago⁶, yet are still able to mate and produce viable (but sterile) offspring. Their compatibility allowed us to identify *cis*- and *trans*-regulatory changes by comparing the regulation of species-specific alleles in a common, hybrid genetic background. Differential expression of two alleles in the same cellular environment indicates functional *cis*-regulatory differences^{4,5,7}; thus, asymmetric allelic expression in F₁ hybrids implies *cis*-regulatory divergence. If *trans*-regulation diverges between species, the collection of *trans*-acting factors in

hybrids will be different from that in one or both of the parental species. As a result, the relative allelic expression in hybrids will differ from the relative gene expression between species. *Trans*-regulatory divergence was inferred for any gene with significant differences in the ratio of species-specific transcripts between F₁ hybrids and the parental species.

To measure the relative abundances of species-specific transcripts, we analysed RNA and DNA, extracted separately from pools of female flies containing either 14 F₁ hybrids or 7 *D. melanogaster* and 7 *D. simulans* individuals (Fig. 1a). Four complementary DNA samples were generated from each RNA extraction and used to measure the relative abundance of species-specific transcripts. At least two replicate hybrid pools (eight cDNA measurements in total) and four replicate parental pools (16 cDNA measurements in total) were analysed for each gene. The ratio of *D. melanogaster* to *D. simulans* alleles was also measured in genomic DNA from each pool in duplicate and used to correct cDNA ratios for allelic differences in extraction and/or amplification (Supplementary Information). For each gene, the ratio of expression between species (designated Mel/Sim) and the ratio of species-specific transcripts in F₁ hybrids (designated Mel_{F1}/Sim_{F1}) were quantified, normalized and averaged across replicate pools.

Pyrosequencing⁸ was used to measure the relative abundance of *D. melanogaster* and *D. simulans* alleles directly in cDNA and DNA samples. A single nucleotide difference that distinguished the species-specific transcripts was identified for each gene, and polymerase chain reaction (PCR) primers were used to amplify a region of DNA containing the divergent site (Fig. 1b). After denaturing the PCR products, an internal primer ('Pyro-reverse' in Fig. 1b) was annealed and extended one base at a time in a Pyrosequencing

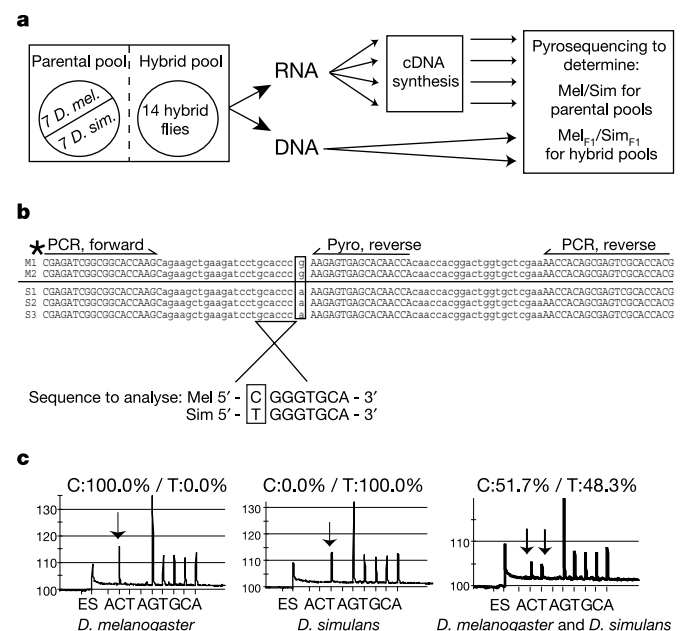


Figure 1 Pyrosequencing⁸ measures allelic gene expression. **a**, Overview of sample collection. **b**, Example assay of CG14770. Aligned coding sequences from *D. melanogaster* (M1, M2) and *D. simulans* (S1–S3) identify transcribed fixed differences (box). The location of PCR and Pyrosequencing primers are shown with an asterisk indicating the biotinylated PCR primer. **c**, Pyrosequencing produces a pyrogram with peaks whose height is directly proportional to the quantity of nucleotide added to the extending Pyrosequencing primer. Pyrograms from reactions using genomic DNA from *D. melanogaster* (left), *D. simulans* (middle), and both species (right) are shown. Arrows indicate peaks reflecting nucleotide incorporation at the divergent site. ES indicates the addition of enzymes and substrate, respectively.

reaction, which couples DNA synthesis with a series of enzymatic reactions that produce stoichiometric quantities of light⁹. The amount of light generated by the incorporation of species-specific bases was directly proportional to the relative abundance of the *D. melanogaster* and *D. simulans* alleles in the PCR template (Fig. 1c, Supplementary Fig. 1).

Thirty-four genes previously shown to have a significant expression difference between *D. melanogaster* and *D. simulans* were selected for analysis¹⁰; of these genes, 29 showed significant expression differences between species in this study, ranging from 1.1-fold to 6.7-fold (Fig. 2a, *t*-test, $H_0: \text{Mel}/\text{Sim} = 1$, $P < 0.05$). Because we examined a different developmental stage in different strains of flies from those in the initial study¹⁰, it is not surprising that we observed distinct expression levels. Comparison with an independent microarray study of expression differences between *D. melanogaster* and *D. simulans* adults¹¹ indicates that this gene set is representative of interspecific expression differences genome-wide (Supplementary Information).

Unequal expression of the *D. melanogaster* and *D. simulans* alleles in a hybrid genetic background indicates differences in *cis*-regulation. Allelic expression of species-specific alleles in F₁ hybrids showed that nearly all (28 of 29) genes with interspecific expression differences had evolved changes in *cis*-regulatory function (Fig. 2b;

t-test, $H_0: \text{Mel}_{F1}/\text{Sim}_{F1} = 1$, $P < 0.05$). Differential expression of species-specific alleles ranged from 1.1-fold to 6.7-fold, with the more highly expressed allele derived from *D. melanogaster* half of the time. Two of the five genes with similar expression between species also showed evidence of *cis*-regulatory divergence (Fig. 2a, b, arrows; *t*-test, $H_0: \text{Mel}_{F1}/\text{Sim}_{F1} = 1$, $P < 0.05$), indicating that compensatory *cis*- and *trans*-regulatory changes have evolved that maintain gene expression. To ensure that differences in allelic expression were caused by *cis*-regulatory changes rather than parent-of-origin effects, we compared the ratio of species-specific transcripts for 24 genes in F₁ hybrids from reciprocal crosses. The direction of cross used to produce the hybrids generally had a negligible effect on the expression levels of the *D. melanogaster* and *D. simulans* alleles (Supplementary Fig. 2).

Cis-regulatory differences seem to be extremely common between orthologous genes in *D. melanogaster* and *D. simulans*. But are changes in *cis*-regulatory function usually the primary cause of interspecific expression differences, or do they contribute relatively little in comparison with genetic changes affecting *trans*-regulation? To address this question, we compared the expression difference between species (Mel/Sim) with the relative abundance of species-specific transcripts in F₁ hybrids (Mel_{F1}/Sim_{F1}; Fig. 3). If *cis*-regulatory divergence completely explains the difference between species, the hybrid and parental expression ratios will be the same (Mel/Sim = Mel_{F1}/Sim_{F1}) and points in Fig. 3 will fall on the diagonal. In contrast, if only *trans*-regulatory differences cause the divergent expression, the *D. melanogaster* and *D. simulans* alleles will be equally expressed in F₁ hybrids (Mel_{F1}/Sim_{F1} = 1) and

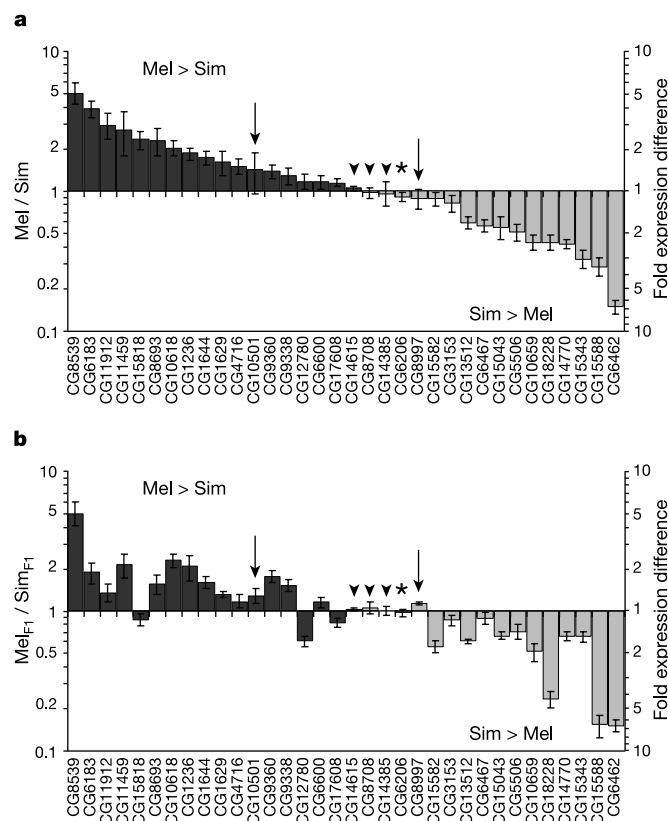


Figure 2 *Cis*-regulatory changes frequently contribute to interspecific expression differences. **a**, Relative gene expression levels between species (Mel/Sim). **b**, Relative allelic expression in F₁ hybrids (Mel_{F1}/Sim_{F1}). The gene order is the same in both panels, with data plotted on a logarithmic scale. The normalized ratio of *D. melanogaster* to *D. simulans* transcripts is shown at the left; the corresponding fold expression changes are shown at the right. Error bars show 95% confidence intervals. Arrows indicate genes with compensatory *cis*- and *trans*-regulatory changes, and arrowheads show genes with conserved *cis*- and *trans*-regulation. Asterisks indicate CG6206, the only gene that showed divergent expression without significant *cis*-regulatory differences.

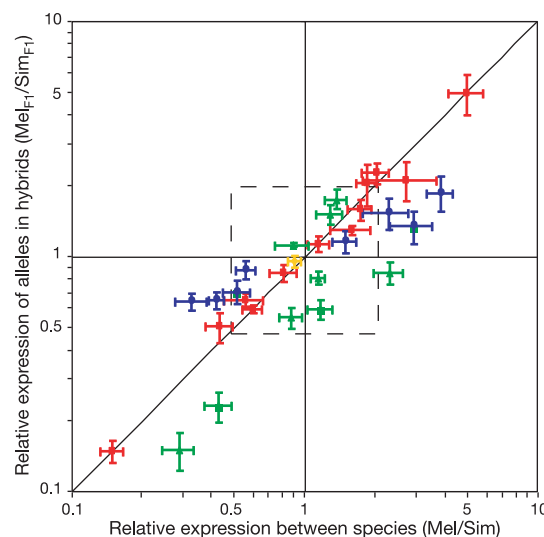


Figure 3 Evolutionary changes in both *cis*- and *trans*-regulation underlie interspecific expression differences. The relative expression between species (Mel/Sim) is plotted against the relative expression of species-specific alleles in hybrids (Mel_{F1}/Sim_{F1}) on a logarithmic scale. The dashed box indicates twofold expression differences. Error bars show 95% confidence intervals. Genes are colour-coded according to our inference of *cis*- and *trans*-regulatory changes: red, *cis*-regulatory changes can completely explain interspecific expression differences; blue, *cis*-regulatory changes account for a fraction of the expression difference, with *trans*-regulatory changes explaining the remaining fraction; green, *cis*- and *trans*-regulatory changes with antagonistic effects have evolved; yellow, CG6206, which displayed no evidence of *cis*- or *trans*-regulatory divergence ($P = 0.09$ and $P = 0.25$, respectively) despite a small significant expression difference between species (Mel/Sim = 0.91, $P = 0.005$). The diagonal line indicates 100% *cis*-regulatory divergence; the bold horizontal line indicates 100% *trans*-regulatory divergence.

points will fall along the horizontal axis. Genes affected by both *cis*- and *trans*-regulatory divergence are indicated in Fig. 3 by points that do not fall on either of these lines.

Cis-regulatory changes could completely explain interspecific expression differences for almost half (12 of 28) of the genes with divergent *cis*-regulatory functions (red in Fig. 3; *t*-test, $H_0: \text{Mel}/\text{Sim} = \text{Mel}_{F1}/\text{Sim}_{F1}$, $P > 0.05$). The remaining 16 genes displayed evidence of both *cis*- and *trans*-regulatory changes ($\text{Mel}_{F1}/\text{Sim}_{F1} \neq 1$ and $\text{Mel}_{F1}/\text{Sim}_{F1} \neq \text{Mel}/\text{Sim}$). For eight genes, differences in *cis*-regulation were less extreme than the interspecific expression difference (blue in Fig. 3). Genetic changes affecting *trans*-acting factors are responsible for the remaining expression difference between species. The other eight genes showed allelic expression differences in F_1 hybrids that were greater than, or in the opposite direction from, the interspecific differences (green in Fig. 3). Evolutionary changes in *cis*- and *trans*-regulation with opposing effects on gene expression must underlie interspecific differences for these genes. Such changes are consistent with the coevolution of *cis*- and *trans*-regulatory factors. Overall, *cis*-regulatory differences affected 97% of genes with divergent expression between *Drosophila* species, with at least 57% also influenced by *trans*-regulatory changes. The presence of *trans*-regulatory differences was not related to the magnitude of the expression difference between species (Kruskal–Wallis test, $P = 0.9$).

The distribution of *cis*- and *trans*-regulatory changes provides insight into the evolution of regulatory networks. To a first approximation, these networks are composed of two classes of gene¹²: regulatory (for example those encoding transcription factors or signalling molecules) and structural (for example those encoding enzymes or cellular components). Regulatory genes (Fig. 4, filled circles) comprise internal connections in the network, and changes in their expression often have *trans*-acting effects on the expression of many other genes. In contrast, structural genes (Fig. 4, open circles) typically lie at terminal nodes of the network where changes in their expression have more limited effects on expression of other genes. On the basis of the available annotations of molecular function^{13,14}, our data set contains primarily structural genes (Supplementary Table 1). Analysis of structural genes provides a complete view of the regulatory network, with gene-specific changes at terminal nodes in the network detected as *cis*-regulatory changes and genetic changes located in upstream regulatory genes identifiable as *trans*-regulatory differences.

Two extreme models can be considered to explain the myriad differences in gene expression between species: a few *cis*-regulatory changes affecting regulatory genes could have widespread *trans*-acting effects on the expression of many downstream genes (Fig. 4a), or each structural gene's expression could be altered by its own *cis*-acting genetic changes (Fig. 4b). We found that interspecific

expression differences were almost always caused, at least in part, by *cis*-regulatory changes in structural genes, and that differences in *trans*-regulation also affected half of the genes (Fig. 4c). The prevalence of *cis*-regulatory changes suggests that differences in gene expression between *D. melanogaster* and *D. simulans* could have evolved by changing the expression of structural genes one gene at a time. However, regulatory differences observed in extant flies might not have been the original source of expression divergence.

With the availability of DNA microarrays, whole genomes can now be compared for differences in gene expression within and between species. By combining expression analysis with genetic methods, such as quantitative trait locus mapping or the approach presented here, the genetic basis of variable gene expression is beginning to be elucidated^{4,5,15–18}. Distinguishing between *cis*- and *trans*-regulatory changes is the first step in this process, and it provides a crucial foothold for identifying the specific nucleotide changes underlying gene expression differences. □

Methods

Fly strains and crosses

A *D. melanogaster* zygotic hybrid rescue (*zhr*) strain (provided by A. Orr) and the Tsimbazaza *D. simulans* strain (provided by H. Hollocher) were crossed to produce F_1 hybrids. These strains reduce premating isolation between *D. melanogaster* and *D. simulans* and rescue the lethality of hybrid females^{19–22}. Female flies from parental lines and F_1 hybrid females from reciprocal interspecific crosses were collected within 4 h of emergence, aged for 24 h, snap-frozen in liquid nitrogen, and stored at -70°C . The *D. melanogaster* In(1)AB strain and *D. simulans* C167.4 and male hybrid rescue (*mhr*) strains (provided by D. Barbash) were also used for DNA sequencing.

Gene selection

Genes were selected on the basis of the microarray data of Rifkin *et al.*¹⁰, which measured changes in gene expression during metamorphosis between *D. melanogaster* and *D. simulans*. Each gene selected showed evidence of lineage-specific selection and had expression in *D. simulans* that was at least twofold different from all four strains of *D. melanogaster* analysed. Because morphological identification of F_1 hybrid females during the larval and pupal stages examined by Rifkin *et al.* is difficult, we analysed adult female flies instead. In addition, because the fly strains used by Rifkin *et al.*¹⁰ do not mate, we used the *zhr* and Tsimbazaza strains.

Extraction of nucleic acids and preparation of cDNA

For each pool of 14 flies, DNA and RNA were extracted separately from a single homogenate by using a modified protocol for the SV Total RNA Isolation System (Promega)²³. The homogenate was passed over a column that retained DNA; the flow-through (containing RNA and proteins) was then passed over a second column that bound RNA. RNA samples were treated with DNase during extraction and immediately before cDNA synthesis. RNA was reverse transcribed into cDNA with the use of a poly(T) primer and standard protocols. PCR reactions with primers annealing to introns confirmed that cDNA samples were free of genomic DNA. Protocols are available from the authors on request.

Allele quantification with Pyrosequencing

Using primer sequences listed in ref. 10, we amplified and sequenced 200–800 base pairs of transcribed DNA from each gene in two *D. melanogaster* and three *D. simulans* strains. After the identification of species-specific nucleotide differences, PCR primers annealing to conserved sequences were used to amplify regions of sequence including the divergent sites. An internal primer, also matching a conserved sequence, was then annealed. Pyrosequencing reactions, performed in accordance with the manufacturer's instructions (www.pyrosequencing.com), were used to measure the relative abundance of the two alleles in genomic DNA and cDNA samples from both parental and hybrid pools. Primer sequences and PCR annealing temperatures are available from the authors on request. Pyrosequencing software reports a peak height directly proportional to the number of molecules incorporated into the growing DNA chain. Custom Perl scripts were used to calculate the ratio of peak heights associated with incorporation of allele-specific bases ($\text{Mel}_{F1}/\text{Sim}_{F1}$ or Mel/Sim), which corresponds to the relative abundance of the *D. melanogaster* and *D. simulans* alleles in the starting sample (Supplementary Fig. 1). cDNA ratios were normalized with genomic DNA measurements as described in Supplementary Information. Because both alleles are extracted and measured in a single sample, this method is insensitive to differences in extraction efficiency and eliminates the need for 'control' genes or quantification of total RNA recovery.

Statistical analysis

After normalization, replicate pools were compared for each gene by using analysis of variance. Significant heterogeneity between pools was observed for 31 of 34 parental pools, but for only 9 of 34 hybrid pools (Supplementary Fig. 3). This variation can be caused either by residual genetic variation within inbred strains or, more probably, by environmental variation introduced during fly rearing. However, because no single pool

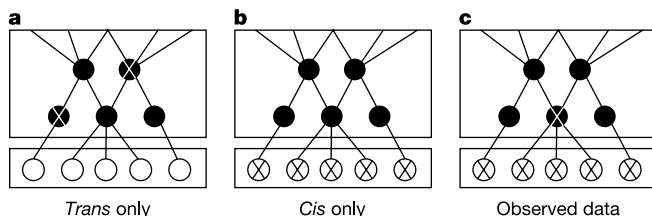


Figure 4 Models of regulatory divergence. A hypothetical regulatory network is shown with filled circles representing regulatory genes, open circles representing structural genes (all with altered expression), and lines representing regulatory interactions. Crosses indicate genes harbouring a regulatory change. **a**, **b**, Two extreme models of regulatory divergence in which either *trans*-regulatory differences (**a**) or *cis*-regulatory differences (**b**) are solely responsible for expression differences. **c**, A distribution of *cis*- and *trans*-regulatory changes consistent with our data.

was a consistent outlier for all genes, data from all pools were included in our analysis. Differences between parental pools increase our confidence intervals for the interspecific expression difference and can reduce our power for detecting *trans*-regulatory differences. To deal with this issue, interspecific expression differences for genes with unusually large variance between the four original pools were measured in an additional three parental pools.

The normalization procedure used to correct cDNA measurements for experimental bias (Supplementary Information) prohibits a standard nested analysis of variance, but a *t*-test provided a simple and robust test of our null hypotheses. Two-tailed *t*-tests were used to identify *cis*-regulatory divergence ($H_0: \text{Mel}_{F1}/\text{Sim}_{F1} = 1$), interspecific expression differences ($H_0: \text{Mel}/\text{Sim} = 1$) and parent-of-origin effects ($H_0: \text{Mel}_{F1}/\text{Sim}_{F1} = \text{Mel}_{F1}/\text{Sim}_{F1}$ in reciprocal crosses). To identify *trans*-regulatory divergence, two-sided *t*-tests (with the Cochran correction for unequal variances) and nonparametric Mann–Whitney *U*-tests were used to compare relative expression between hybrid and parental pools. The decision to accept or reject the null hypothesis ($H_0: \text{Mel}_{F1}/\text{Sim}_{F1} = \text{Mel}/\text{Sim}$) was the same for both tests for all except three genes, and *t*-test significance was ultimately used to infer *trans*-regulatory divergence. All statistical analyses were performed with SAS software v. 8.2 (SAS Institute, Cary, North Carolina) and are shown in Supplementary Table 2.

Received 29 February; accepted 1 June 2004; doi:10.1038/nature02698.

- Davidson, E. H. *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego, 2001).
- Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Oxford, 2001).
- Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
- Cowles, C. R., Hirschhorn, J. N., Althuler, D. & Lander, E. S. Detection of regulatory variation in mouse genes. *Nature Genet.* **32**, 432–437 (2002).
- Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297**, 1143 (2002).
- Powell, J. R. *Progress and Prospects in Evolutionary Biology: The Drosophila Model* (Oxford Univ. Press, New York, 1997).
- Pastinen, T. *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiol. Genomics* **16**, 184–193 (2004).
- Ahmadian, A. *et al.* Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**, 103–110 (2000).
- Neve, B. *et al.* Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. *Biotechniques* **32**, 1138–1142 (2002).
- Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33**, 138–144 (2003).
- Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D. & Hartl, D. L. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745 (2003).
- Davidson, E. H., McClay, D. R. & Hood, L. Regulatory gene networks and the properties of the developmental process. *Proc. Natl Acad. Sci. USA* **100**, 1475–1480.
- Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
- The FlyBase Consortium, The FlyBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **31**, 172–175 (2003).
- Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35**, 57–64 (2003).
- Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
- Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
- Montooth, K. L., Marden, J. H. & Clark, A. G. Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. *Genetics* **165**, 623–635 (2003).
- Davis, A. W. *et al.* Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans*. *Nature* **380**, 157–159 (1996).
- Sawamura, K., Yamamoto, M. T. & Watanabe, T. K. Hybrid lethal systems in the *Drosophila melanogaster* species complex. II. The *Zygotic hybrid rescue* (*Zhr*) gene of *D. melanogaster*. *Genetics* **133**, 307–313 (1993).
- Hollocher, H., Agopian, K., Waterbury, J., O'Neill, R. W. & Davis, A. W. Characterization of defects in adult germline development and oogenesis of sterile and rescued female hybrids in crosses between *Drosophila simulans* and *Drosophila melanogaster*. *J. Exp. Zool.* **288**, 205–218 (2000).
- Barbash, D. A., Roote, J. & Ashburner, M. The *Drosophila melanogaster* hybrid male rescue gene causes inviability in male and female species hybrids. *Genetics* **154**, 1747–1771 (2000).
- Otto, P., Kephart, D., Bitner, R., Huber, S. & Volkerding, K. Separate isolation of genomic DNA and total RNA from single samples using the SV Total RNA Isolation System. *Promega Notes* **69**, 19–24 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Kumari and S. Madhavarapu for experimental assistance, K. Montooth for statistical advice, and K. Montooth, B. Payseur, A. Fiumera, T. Schlenke and E. Hill for comments on the manuscript. Funding for this project was provided by NIH grants to A.G.C. P.J.W. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation, and B.K.H. was funded by a Howard Hughes Undergraduate Research award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.W. (pw72@cornell.edu).

Evidence for dynamically organized modularity in the yeast protein–protein interaction network

Jing-Dong J. Han¹, Nicolas Bertin¹, Tong Hao¹, Debra S. Goldberg², Gabriel F. Berriz², Lan V. Zhang², Denis Dupuy¹, Albertha J. M. Walhout^{1*}, Michael E. Cusick¹, Frederick P. Roth² & Marc Vidal¹

¹Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

* Present address: Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

In apparently scale-free protein–protein interaction networks, or ‘interactome’ networks^{1,2}, most proteins interact with few partners, whereas a small but significant proportion of proteins, the ‘hubs’, interact with many partners. Both biological and non-biological scale-free networks are particularly resistant to random node removal but are extremely sensitive to the targeted removal of hubs¹. A link between the potential scale-free topology of interactome networks and genetic robustness^{3,4} seems to exist, because knockouts of yeast genes^{5,6} encoding hubs are approximately threefold more likely to confer lethality than those of non-hubs¹. Here we investigate how hubs might contribute to robustness and other cellular properties for protein–protein interactions dynamically regulated both in time and in space. We uncovered two types of hub: ‘party’ hubs, which interact with most of their partners simultaneously, and ‘date’ hubs, which bind their different partners at different times or locations. Both *in silico* studies of network connectivity and genetic interactions described *in vivo* support a model of organized modularity in which date hubs organize the proteome, connecting biological processes—or modules⁷—to each other, whereas party hubs function inside modules.

The biological role of topological hubs, so far considered in static representations of interactome networks without information on the functional states of these networks—that is, dynamic or steady state⁸—might vary depending on the timing and location of the interactions they mediate (Fig. 1a). Because accurate temporal parameters are not yet available for many protein–protein interactions, we estimated temporal characteristics of hubs and their partners by using compilations of yeast messenger RNA expression profiling data⁹.

Hubs connected by false-positive interactions¹⁰ would be uncorrelated in mRNA expression with their interaction partners^{9,11}, and would resemble date hubs. To minimize false positives, we first generated a high-quality yeast interaction data set by intersecting data generated by several different interaction detection methods (see Methods). The resulting ‘filtered yeast interactome’ (FYI) data set contains 2,493 high-confidence interactions, each observed by at least two different methods (Supplementary Fig. 1). FYI is a high-quality network enriched for genuine positives (Supplementary Information and Supplementary Fig. 2). The FYI network contains 1,379 proteins with an average degree of 3.6 interactions per protein and a large connected component of 778 proteins. Its degree distribution follows the power law that characterizes scale-free networks (Supplementary Fig. 3). FYI hubs were characterized with an expression-profiling compendium of 315 data points for most yeast genes across five different experimental conditions (referred to below as the ‘yeast expression compendium’⁹). For each hub we calculated the average of Pearson

correlation coefficients between the hub and each of its respective partners for mRNA expression (see Methods). Strikingly, the average PCCs of hubs, defined as nodes (proteins) with degree k greater than 5, follow a bimodal distribution in the whole compendium (Fig. 1b, red curve). In contrast, the average PCCs of non-hubs, defined as nodes with degree k of 5 or less, show a normal distribution centred on 0.1 (Fig. 1b, cyan curve, and Supplementary Fig. 4). In randomized interactome networks of the same topology (Supplementary Methods), the average PCCs of hubs also show a normal distribution centred on 0 (Fig. 1b, black curve). This bimodal distribution suggests that hubs can be split into two distinct populations: one with relatively high average PCCs (party hubs) and the other with relatively low average PCCs (date hubs) (Supplementary Information). Furthermore, the bimodal distribution suggests a natural boundary for separating or partitioning date hubs from party hubs.

Party and date hubs were analysed for each individual condition of the yeast expression compendium. Average PCCs of FYI hubs show a clear bimodal distribution for two conditions: 'stress response' and 'cell cycle' (containing 174 and 77 data points, respectively) (Fig. 1b). The three remaining conditions ('phero-

mone treatment', 'sporulation' and 'unfolded protein response') contain fewer data points (45, 10 and 9, respectively), which may explain the absence of a clear bimodal distribution (Fig. 1b). For our subsequent analyses, party hubs are those with an average PCC higher than the threshold indicated by the arrow, in at least one of the five conditions in Fig. 1b (exact cutoffs used in Methods and Supplementary Information). All other hubs were defined as date hubs. Using these criteria, we found 91 date hubs and 108 party hubs in FYI (Supplementary Table 1) after excluding ribosomal hub proteins (Supplementary Information).

Highlighting the biological significance of our date/party hub partitioning is the fact that average PCC values correctly predict the expected date versus party behaviour for several well-characterized protein hubs (Supplementary Table 1 and Supplementary Information). The dynamics of interactome networks should be considered not only by expression timing but also spatial distribution—that is, subcellular localization. We estimated the localization diversity of partners of hubs by using a proteome-wide cellular localization data set¹². Partners of date hubs are significantly more diverse in spatial distribution than partners of party hubs (protein localization diversity evaluated by entropy

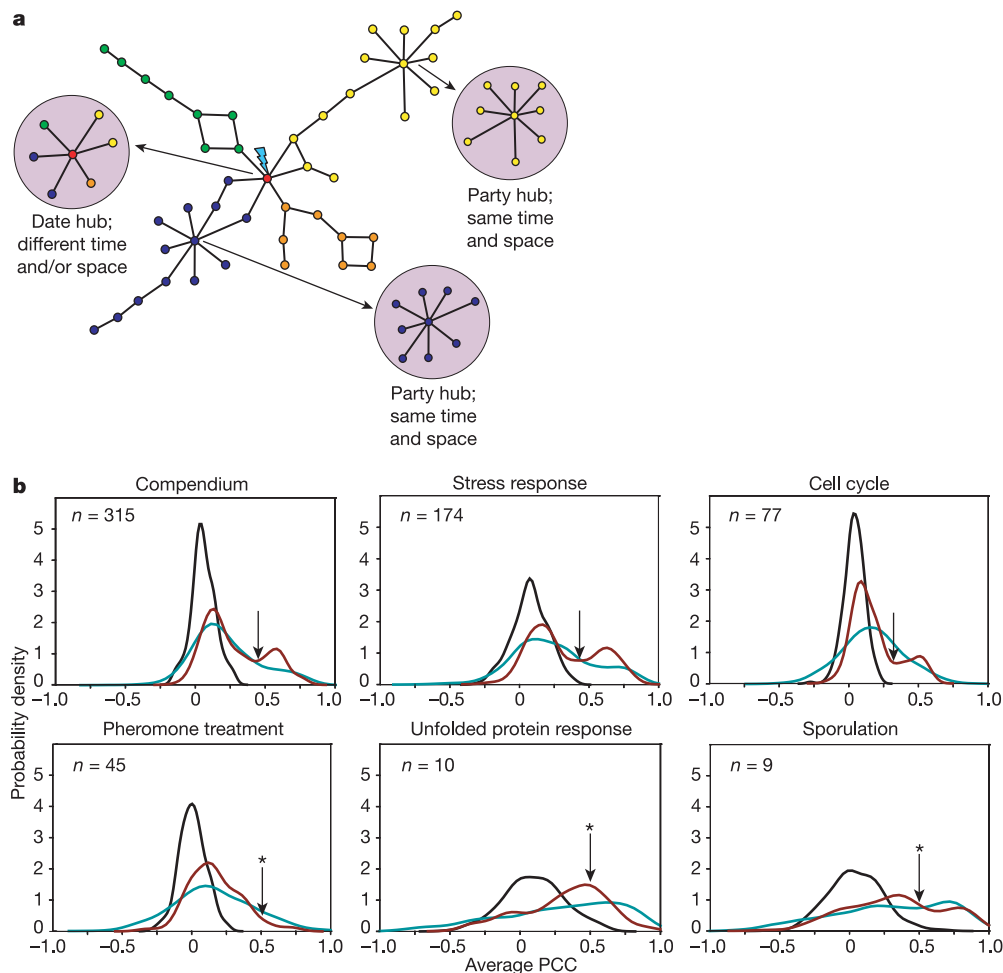


Figure 1 Date and party hubs. **a**, In this schematic protein interaction network, proteins are coloured according to mutual similarity in their mRNA expression patterns. 'Party' hubs are highly correlated in expression with their partners, and presumably interact with them at similar times. The partners of 'date' hubs exhibit more limited co-expression, and presumably the corresponding physical interactions occur at different times and/or different locations. **b**, Probability densities (Supplementary Methods) of the average PCCs were calculated from a global expression profiling compendium⁹ (top left panel). Average PCCs were also independently calculated for each condition constituting the

compendium. The number n in each panel refers to the number of data points for each gene for each condition. Average PCCs for hubs in the FYI (red curve) show a clear bimodal distribution that is used to separate date and party hubs (located by the arrow) for the conditions shown in the top panels. For the conditions in the bottom panels that do not show a clear bimodal distribution, an arbitrary average PCC cutoff of 0.5 was used (details in Methods and Supplementary Information). No bimodal distribution is observed with the average PCCs of non-hub proteins (cyan curve) or for hubs in randomized networks (black curve).

calculation; Student's t -test $P < 0.05$). Hence, the distinction between date and party hubs obtained from gene expression is recapitulated by protein localization data.

When removed from the interactome network, party and date hubs have distinct effects on the overall topology. We used an *in silico* strategy¹³ that simulates the effect of specifically removing (attacking) hubs in the FYI network on the characteristic path length of the main component of the network. The characteristic path length, defined as the average distance (shortest path length) between node pairs, reflects the overall network connectivity¹³. As expected, successive attacks against FYI hubs, starting from the most con-

nected hubs, without distinguishing between party and date hubs, have a significantly more deleterious effect on the network integrity than the removal of random proteins (failure)¹³ (Fig. 2a, b). However, this *in silico* experiment revealed an unexpected and striking difference between party and date hubs. Removal of party hubs does not affect connectivity and thus resembles failures (Fig. 2a, b), whereas attacks directed against date hubs account for a vast majority of the effect observed when attacking all hubs (Fig. 2a, b).

To rule out the possibility that small variations of local topology cause the differences observed above, we performed additional

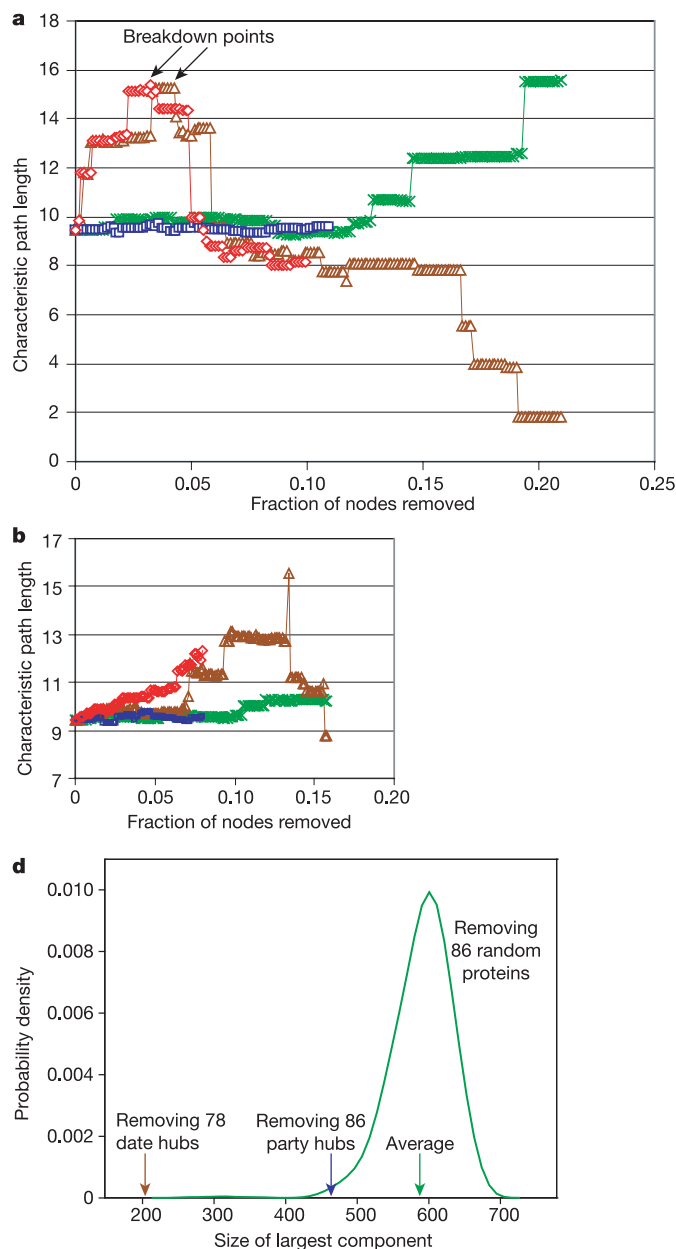


Figure 2 Date hubs are central to network topology. **a**, The effects on the characteristic path length of the network on gradual node removal. Random removal of nodes ('failures') is represented by the green line, attacks against all hubs by the brown line, attacks against party hubs by the blue line, and attacks against date hubs by the red line. The 'breakdown point' is the threshold after which the main component of the network starts disintegrating. **b**, Subsets of date and party hubs with comparable degree (k) and clustering coefficients (C_v)¹⁸ were selected for attack (lines are coloured as in **a**). **c**, The main component of the FYI network (top panel) splits into small subnetworks (middle

panel) after the removal of date hubs, whereas it stays almost intact after the removal of party hubs (bottom panel). **d**, The sizes of the largest remaining component after removing all 78 date hubs (red arrow), all 86 party hubs (blue arrow) or 86 randomly selected proteins (green curve). This last experiment has been repeated 1,000 times to determine the empirical P values of recovering sizes similar to those obtained upon removal of date and party hubs (empirical P values are equal to 6×10^{-3} and less than 10^{-3} , respectively).

simulations by attacking FYI with subsets of date and party hubs that show comparable values of C_v (clustering coefficient, a measure of neighbourhood density) and k (number of interaction partners) (see Methods). The differences between date and party hubs were similar to those noted above (Fig. 2b and Supplementary Information). Thus, date and party hubs have markedly different global properties in the interactome network.

The main component that remains after the removal of party hubs is significantly larger than that remaining after the removal of date hubs (Fig. 2c, d, and Supplementary Fig. 5). Conversely, the subnetworks released by date hub removals tend to be larger in size and number than those obtained by party hub removals (Fig. 2c). To test whether FYI subnetworks obtained after the complete removal of date hubs corresponds to small interaction maps of specific biological processes, or modules⁷, we estimated their functional homogeneity by using annotations from the Munich Information Center for Protein Sequences (MIPS) database¹⁴. In comparison with control networks of the same size distribution, most FYI subnetworks were more homogeneous in function (Fig. 3a, Supplementary Information). We could assign a 'most likely' function for each subnetwork by determining the most enriched function category among all nodes over the entire FYI data set (Supplementary Table 2). Thus, subnetworks derived from the FYI by using a definition based on non-biased functional states, and not biased by topology alone, often correspond to known biological modules.

Subnetworks represent not only stable molecular machines or complexes (for example the ribosomal RNA synthesis complex) but also more loosely connected regulatory pathways (for example osmosensing). Indeed, subnetworks have a broad range of average values of PCCs between all protein pairs involved (Fig. 3b). Protein pairs inside subnetworks corresponding to protein complexes tend to show high PCC values, whereas less densely connected regulatory pathway modules tend to show lower PCC values (Fig. 3b).

These results support a model of organized modularity for the yeast proteome, as illustrated when date hubs are reconnected to the modular subnetworks (module compositions at <http://vidal.dfci.harvard.edu/fyi/moduleNet.pl>; Fig. 4a). In this model, date hubs represent global, or 'higher level'¹³, connectors between modules, and party hubs function inside modules, at a 'lower level'¹⁵ of the organization of the proteome. For example, the date hub calmodulin (Cmd1) connects four different biological modules, 'homeostasis of cations', 'protein folding and stabilization', 'budding, cell polarity and filament formation' and 'endoplasmic reticulum', whereas the party hubs Sec17, Sec 22 and Vti1 all function within the 'endoplasmic reticulum' module (Fig. 4a, inset).

The organized modularity model predicts that experimental perturbations of date hubs *in vivo* should confer different effects from perturbations of party hubs. In single-gene knockout experiments^{5,6}, similar proportions of party and date hubs score as essential (Fig. 4b). Although party hubs tend to mediate their role locally within modules, they can still mediate unique functions in essential modules and thus score as essential genes, explaining the similar essentiality rate between date and party hubs.

In contrast, genetic perturbations of date hubs tend to sensitize the proteome to other perturbations, more so than perturbations of party hubs (Fig. 4c). Among all genetic interactions published by individual laboratories and curated in MIPS¹⁴, genetic interactions involving date hubs are twice as prevalent as those involving party hubs or only non-hub proteins ($P < 10^{-5}$) (Fig. 4c and Supplementary Methods). Assuming that date hubs are not more likely to be studied than party hubs, the higher rate of observed genetic interactions for date hubs suggests that they have a central role in organizing the modularity of the yeast proteome. Conversely, the lower rate of observed genetic interactions for party hubs reflects their localized role within isolated, encapsulated regions of the interactome.

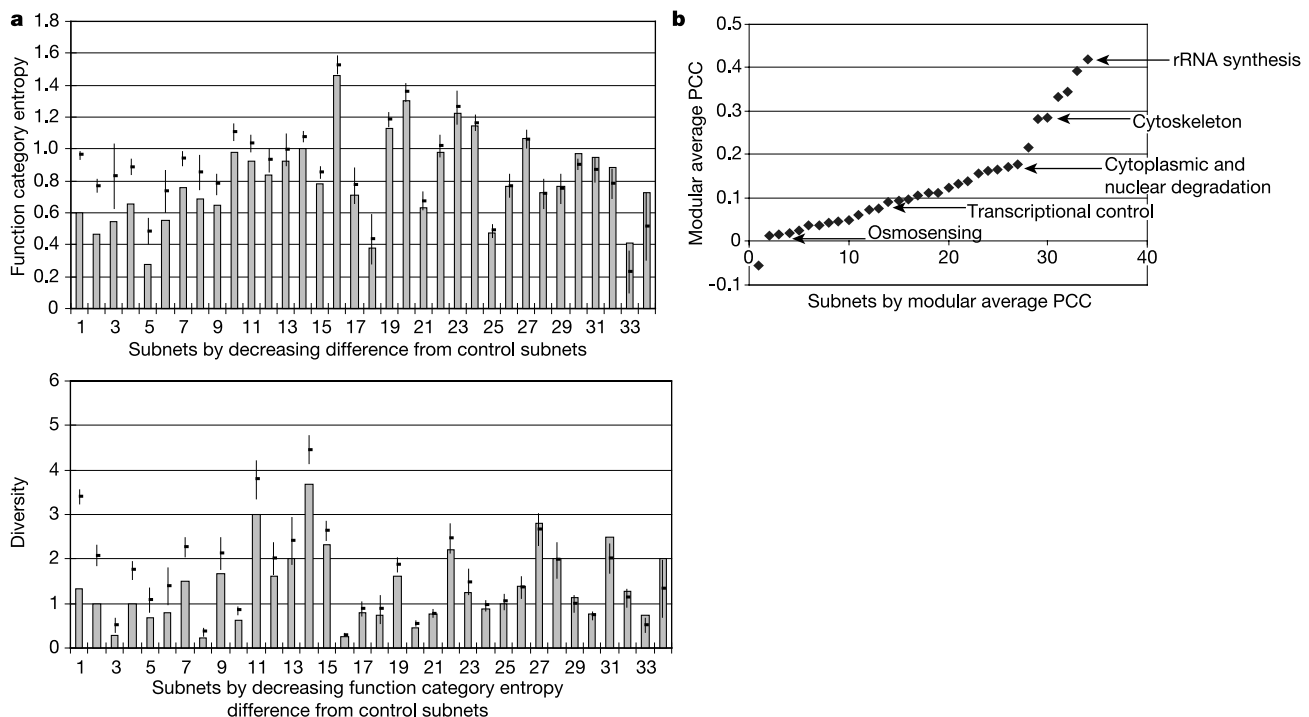


Figure 3 Properties of subnetworks. **a**, Most subnetworks generated by removing date hubs represent functionally homogeneous modules. Grey bars represent the function category entropy (upper panel) or diversity (lower panel) of the subnets revealed by removing date hubs from the original largest component. Lower entropy means greater homogeneity. Dots are the average function category entropy or diversity values of 200

control subnetworks; lines mark one standard deviation on each side of the average.

b, Subnetworks are probably both complexes and more loosely connected modules. PCC values between all genes within each module are plotted against rank on increasing PCCs. The arrows indicate several examples.



interactions obtained by using a 'matrix' representation; that is, all pairwise interactions between all components of a complex), and last, the MIPS physical interactions list (excluding genome-scale experiments: 1,285 interactions).

Transcriptome profiling data set and average PCC calculation

The 'conditions' expression profile compendium was obtained from ref. 9. For average PCC calculation over all profiles, each of the five condition data sets in this compendium was normalized with Z-score normalization¹⁰; that is, the expression measurement for each gene was adjusted to have a mean of 0 and a standard deviation of 1 across all conditions. For the calculation of average PCC within each condition, the original log₂ fold change values were used, with or without filtering for genes that displayed at least 1.5-fold changes.

Average PCC cutoff to divide date and party hubs

For those conditions that showed a bimodal distribution, we selected the average PCC cutoff at the valley between the two peaks. For those conditions that did not show a clear division, we used an arbitrary cutoff of 0.5, a value slightly higher than those corresponding to bimodal distributions. In addition, we tested an alternative method, which yielded 104 party hubs (Supplementary Information). The data based on the second partition are very similar to those presented in the main text (data not shown). Thus, our partitioning strategy is tolerant to a small fraction of spillover between the date and party hubs.

Attacking hubs with comparable clustering coefficients and degrees

The difference between local clustering density or degree distribution of the date and party hubs could explain their different behaviour in our *in silico* network attacks. To refute this hypothesis, we calculated C_v for each hub as described²⁷ and subsequently selected a subset of 62 date and 62 party hubs in the largest component of comparable k and C_v by first removing the 8 date hubs with the highest k and the 16 party hubs with the lowest k , and subsequently removing the 8 date hubs with the lowest C_v and the 8 party hubs with the highest C_v . This resulted in sets of date and party hubs with nearly identical average values of k or C_v (Wilcoxon rank sum test $P > 0.5$ for both; χ^2 test $P = 0.72$ and 0.37 for C_v and k , respectively, based on counts at either side of their mean values).

Calculations

Localization entropy of partners of a hub. This was calculated as $-\sum L_i \log L_i$ (ref. 28), where L_i is the frequency of appearance of a subcellular localization i . $L_i = T_i / \sum_i T_i$, where T_i is the number of times that the subcellular localization i associated with all partners of a hub and n is the number of distinct subcellular localizations associated with all partners of the hub. A large-scale localization data set¹³, excluding the broad high-level categories 'cytoplasm' and 'nucleus', was used.

Function category entropy of a subnet. This was calculated as $-\sum F_i \log F_i$ (ref. 28), where F_i is the frequency of appearance of a function category i . $F_i = T_i / \sum_i T_i$, where T_i is the number of times that the function category i appears in the subnetwork and n is the number of distinct function categories present in the subnetwork. MIPS function categories, excluding the broad high-level categories 'cytoplasm', 'mitochondrion' and 'nucleus', were used.

Function diversity of a subnet. This was calculated as the number of unique function categories of a subnetwork divided by the total number of members in the subnetwork. It measures the number of unique function categories per member and hence serves as an intuitive way of determining functional diversity.

Genetic interaction density. This was calculated as the ratio of the actual number of interactions found divided by the potential number of interactions¹¹. A detailed description can be found in Supplementary Methods.

Received 16 December 2003; accepted 6 April 2004; doi:10.1038/nature02555.

- Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
- Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
- Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355–361 (2000).
- Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
- Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
- Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. & Palsson, B. O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28**, 250–258 (2003).
- Kemmeren, P. *et al.* Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* **9**, 1133–1143 (2002).
- von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
- Ge, H., Liu, Z., Church, G. M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nature Genet.* **29**, 482–486 (2001).
- Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Albert, R., Jeong, H. & Barabasi, A. L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
- Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30**, 31–34 (2002).

- Ideker, T. & Lauffenburger, D. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.* **21**, 255–262 (2003).
- Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
- Le Pont, F. *et al.* A new scale for measuring dynamic patterns of sexual partnership and concurrency: application to three French Caribbean regions. *Sex. Transm. Dis.* **30**, 6–9 (2003).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
- Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**, 277–282 (1997).
- Fromont-Racine, M. *et al.* Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast* **17**, 95–110 (2000).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**, 248–250 (2003).
- Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
- Dandekar, T., Snel, B., Huynen, M. & Bork, P. Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* **23**, 324–328 (1998).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
- Snel, B., Bork, P. & Huynen, M. A. The identification of functional modules from the genomic association of genes. *Proc. Natl Acad. Sci. USA* **99**, 5890–5895 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank DFCI Research Computing, especially L. Cai and M. Temple, for technical support and computation resources; members of the Vidal laboratory and J. Dekker for suggestions; T. Clingingsmith for administrative assistance; and H. Ge, C. Armstrong, D. Hill, M. Boxem and P.-O. Vidalain for reading the manuscript. F.P.R., G.E.B., L.V.Z. and D.S.G. were supported in part by an institutional grant from the HHMI Biomedical Research Support Program for Medical Schools. L.V.Z. and D.S.G. were supported by a Fu Fellowship and an NSF Postdoctoral Fellowship in Interdisciplinary Informatics, respectively. This work was supported by grants from the NHGRI, NIGMS and NCI awarded to M.V.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.V. (marc_vidal@dfci.harvard.edu).

PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates

Xiaowei Lu¹, Annette G. M. Borchers², Christine Jolicoeur¹, Helen Rayburn¹, Julie C. Baker² & Marc Tessier-Lavigne^{1*}

¹Department of Biological Sciences, Howard Hughes Medical Institute,

²Department of Genetics, Stanford University, Stanford, California 94305, USA

* Present address: Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

In addition to the apical–basal polarity pathway operating in epithelial cells, a planar cell polarity (PCP) pathway establishes polarity within the plane of epithelial tissues and is conserved from *Drosophila* to mammals. In *Drosophila*, a 'core' group of PCP genes including *frizzled* (*fz*), *flamingo/starry night*, *dishevelled* (*dsh*), *Van Gogh/strabismus* and *prickle*, function to regulate wing hair, bristle and ommatidial polarity^{1–3}. In vertebrates, the PCP pathway regulates convergent extension movements and neural tube closure^{3–5}, as well as the orientation of stereociliary bundles of sensory hair cells in the inner ear⁶. Here we show that a mutation in the mouse *protein tyrosine kinase 7* (*PTK7*) gene, which encodes an evolutionarily conserved transmembrane protein with tyrosine kinase homology, disrupts neural tube closure and stereociliary bundle orientation, and shows genetic interactions with a mutation in the mouse *Van Gogh* homologue

vangl2. We also show that PTK7 is dynamically localized during hair cell polarization, and that the *Xenopus* homologue of PTK7 is required for neural convergent extension and neural tube closure. These results identify PTK7 as a novel regulator of PCP in vertebrates.

To identify transmembrane proteins important for neural development, we used a gene trapping⁷ screen and isolated an insertion in the second intron of the mouse *PTK7* gene⁸ (also known as colon carcinoma kinase-4, CCK-4) (ref. 9) (Fig. 1a). This insertion is predicted to result in the fusion of the first 114 amino acids of PTK7 in-frame to a transmembrane β geo (β -galactosidase and neomycin phosphotransferase fusion) construct. The full-length PTK7 protein is predicted to contain a signal peptide, seven immunoglobulin domains, a single transmembrane domain, followed by a cytoplasmic domain with tyrosine kinase homology (Fig. 1b). By northern blot analysis, no wild-type transcript was detected in the mutant; instead a transcript with the expected size of the fused bicistronic RNA was produced (Fig. 1c). Western blot analysis using antibodies generated against the ecto-domain of PTK7 detected two polypeptides in wild-type brain lysates, both of which are absent in the mutant; a polypeptide of the size of the predicted PTK7: β geo

fusion product was detected instead (Fig. 1d). Transmembrane β geo fusion proteins do not reach the cell surface and are retained in intracellular compartments^{7,10}; thus we predict that this insertion results in a null or severely hypomorphic mutation in *PTK7*.

PTK7 mutants die perinatally and display craniorachischisis, a severe form of neural tube defect (NTD) in which the neural tube fails to close from the midbrain–hindbrain boundary to the base of the spine (Fig. 1f). Neural tube closure normally initiates at three sites along the rostral–caudal axis, the first of which (closure 1) occurs in the future cervical region at E8.5 in the mouse. Closures 2 and 3 occur rostral to closure 1, and the spinal cord is closed by caudal ‘spreading’ of closure 1 (ref. 11). In *PTK7* mutants, cranial closure (closures 2 and 3) was not affected (Fig. 1f) but closure 1 failed to initiate, accompanied by a broader floor plate (arrow in Fig. 1j) and wider somites with shorter spacing (bracket in Fig. 1j) compared with wild-type controls. We examined the expression pattern of PTK7 during normal neurulation as indicated by the β geo reporter in heterozygous embryos. PTK7: β geo was dynamically expressed during gastrulation and neurulation, with higher levels in the posterior region of the embryo (Supplementary Fig. 1). This expression pattern was consistent with a role for PTK7 in neural tube closure.

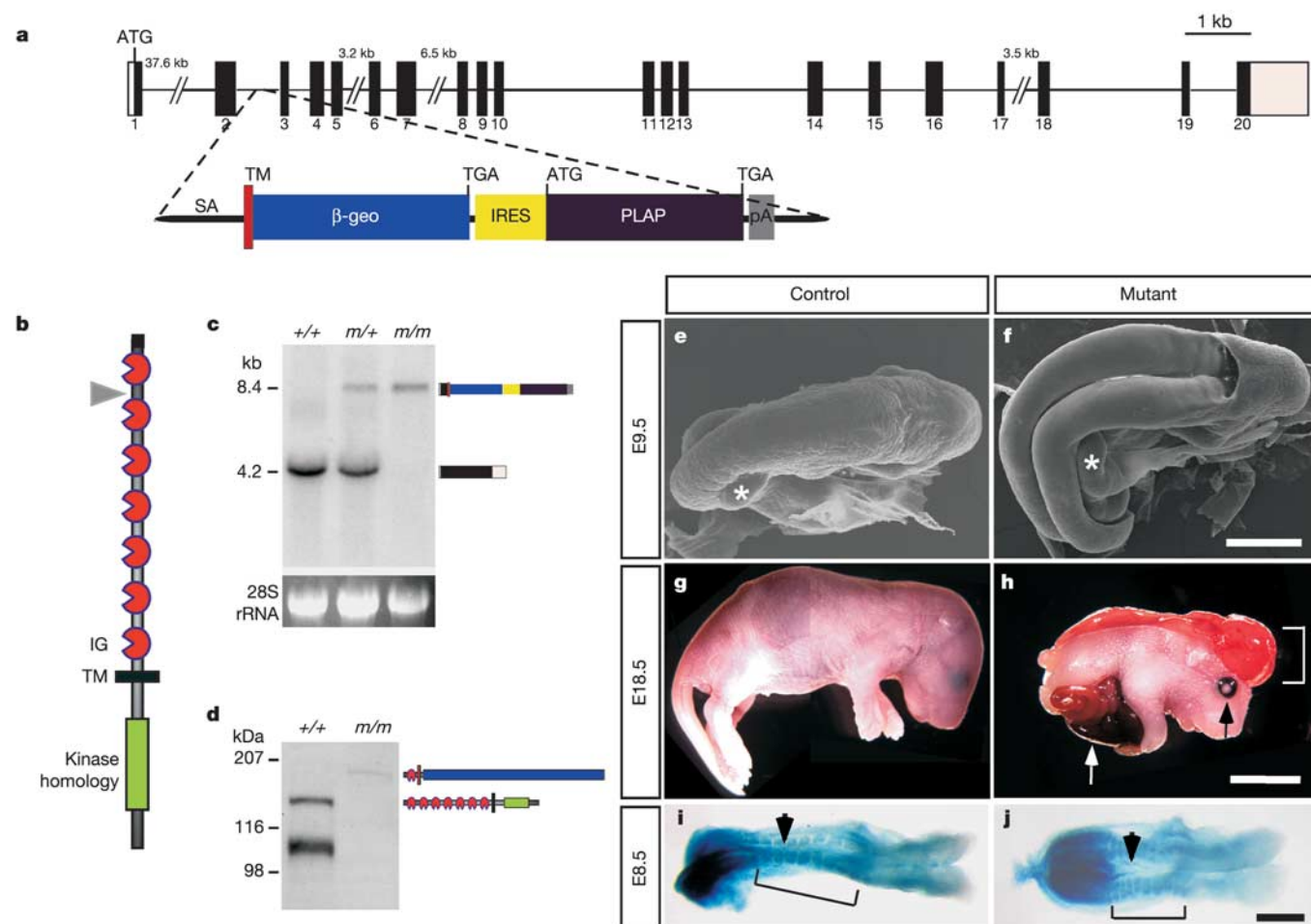


Figure 1 PTK7 loss-of-function phenotypes. **a**, Gene trap insertion into the *PTK7* locus. **b**, Domain structure of PTK7. Arrowhead demarcates gene trap insertion site. TM, transmembrane domain; IG, immunoglobulin domains. **c**, northern blot of total RNA from E16.5 brains; bicistronic and wild-type transcripts are illustrated. **d**, western blot of E15.5 brain lysates. **e–j**, comparison of control (**e**, **g**, **i**) and *PTK7* mutant (**f**, **h**, **j**) embryos, visualized by scanning electron microscopy at E9.5 (**e**, **f**), asterisks indicate forelimb; by light microscopy at E18.5 (**g**, **h**); and by X-gal histochemistry at E8.5 (**i**, **j**). Anterior is to the right. In addition to craniorachischisis (**f** and **h**), which is 100% penetrant, some mutants

exhibit other developmental abnormalities, including hindlimb polydactyly and smaller kidneys (Supplementary Fig. 3), malformed forebrains (bracket in **h**), open eye lids (black arrow in **h**, also observed in *Crc* and *Lp* mutants^{17,24}) and gastroschisis (externalization of abdominal contents, white arrow in **h**, also observed in *Crc* and, infrequently, *Lp* mutants¹⁷). **i**, **j**, at E8.5 (10 somites), neural tube closure fails to initiate in the mutant (**j**), accompanied by a broader floor plate (arrow). Brackets indicate somites. Scale bars: **e**, **f**, 500 μ m; **g**, **h**, 5 mm; **i**, **j**, 200 μ m.

Severe NTD can result from defective dorsal–ventral patterning of the spinal cord, as occurs, for example, when sonic hedgehog signalling is dysregulated¹². However, the expression patterns of several regional markers in the spinal cord of *PTK7* mutants were comparable to those in control animals, suggesting that the NTD in *PTK7* mutants was not caused by a dorsal–ventral patterning defect (Supplementary Fig. 2).

Craniorachischisis has been associated with mutations in a number of mouse orthologues of the *Drosophila* PCP genes. These include *vangl2* (a homologue of *Van Gogh*, mutated in *looptail* mice), *dishevelled-1* and *-2* (ref. 13) and *celsr1* (ref. 14) (a homologue of *flamingo*). Another mouse mutant with craniorachischisis is *circletail* (*Crc*), which carries a mutation in *mScrib1*, a homologue of the *Drosophila* apical–basal polarity gene *scribble*¹⁵. *Celsr1* (ref. 14), *vangl2* and *mScrib1* (ref. 16) are also required to coordinate the orientation of the actin-based stereociliary bundles of sensory hair cells in the organ of Corti, which is the sensory organ for hearing. Therefore we tested whether *PTK7* was also required for stereociliary bundle orientation. Normally, every stereociliary bundle is oriented towards the abneural side of the sensory epithelium. When we examined *PTK7* mutant cochleae at E18.5 by scanning electron microscopy, we found that bundle orientation was disrupted (Fig. 2a–h). In the basal and middle turns of the mutant

cochleae, the regular organization of one row of inner hair cells (IHC) and three rows of outer hair cells (OHCs) separated by pillar cells was largely normal, with occasional and mild disruptions (dashed box in Fig. 2b). Bundle misorientation was most pronounced in the third row of OHCs (OHC3), with some bundles rotated almost 180° from the wild-type position (arrowheads in Fig. 2b, d), whereas IHC and OHC1 and 2 were not affected. Overall, the hair bundle phenotype of *PTK7* mutants was very similar to that of the *Crc* mutants¹⁶. In the apical turn of the *PTK7* mutant cochleae, bundle misorientation appeared more severe, also affecting the IHCs (arrowheads in Fig. 2f). Furthermore, a fourth row of OHCs was observed, and extra IHCs were found scattered outside the row of IHC (asterisk in Fig. 2f). Bundle polarity within individual hair cells was still present: the stereocilia were always organized in a staircase pattern around the tubulin-based kinocilium. Apical–basal polarity of the sensory hair cells appeared normal in *PTK7* mutants, as indicated by apparently normal morphology and expression of E-cadherin (a marker for adherens junctions) (data not shown).

We also examined mutant cochleae at E16.5, a time at which the kinocilium in maturing hair cells moves away from the centre to the peripheral edge of the cell (a first indication of PCP). In IHCs and OHC1 and 2, kinocilium positions were clearly biased towards the

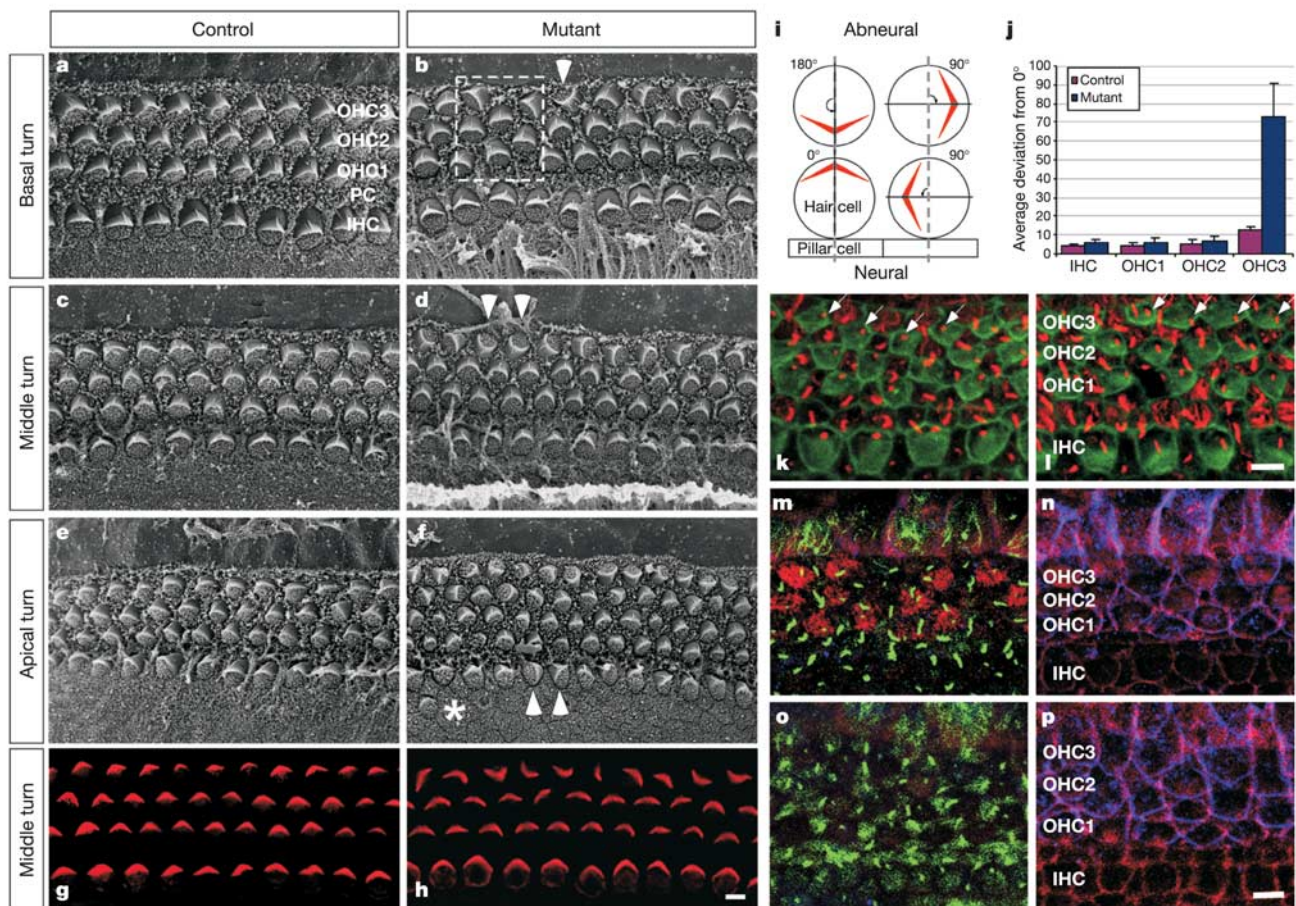


Figure 2 PCP phenotype in *PTK7* mutant cochleae. **a–f**, scanning electron microscopy images of the organ of Corti (positions shown on left) at E18.5. Stereociliary bundle misorientation was most pronounced in OHC3. Dashed box in **b** indicates a mild disruption of the hair cell rows. Arrowheads indicate misoriented bundles. Asterisk in **f** indicates an extra IHC. **g, h**, phalloidin staining of E18.5 organ of Corti showing the same bundle misorientation defect. **i**, Determination of bundle orientation. **j**, Quantification of bundle misorientation based on phalloidin staining. **k, l**, E16.5 control (**k**) and mutant (**l**)

cochleae. The kinocilia (red) are frequently mispositioned in mutant OHC3s (white arrows). F-actin (green). **m–p**, *PTK7* (red) localization in E16.5 wild-type Organ of Corti. Tubulin (green); E-cadherin (blue). **m** and **o** are from one z plane (apical surfaces), and **n** and **p** from one basal in relation to **m** and **o** (adherens junctions). **m** and **n** are developmentally less advanced than **o** and **p**. Note transient apical accumulation of *PTK7* in outer hair cells (compare **m** and **o**). Scale bars, **a–h** and **k–p**, 2 μm. IHC, inner hair cell; OHC, outer hair cell; PC, pillar cell.

abneural side of the cell, as is the case for wild-type hair cells. By contrast, kinocilia in OHC3 in the mutant were positioned randomly, often towards the neural side of the cell (arrows in Fig. 2l), suggesting that PTK7 regulates the establishment of PCP in hair cells. Immunostaining using PTK7-specific antibodies showed dynamic expression of PTK7 during this critical period of PCP signalling. In the wild-type, PTK7 is upregulated in the organ of Corti as differentiation progresses in a wave along both the neural-to-abneural and basal-to-apical axes (data not shown). Interestingly, over the period when the kinocilium moves from the centre to the edge of the cell, PTK7 appears to accumulate transiently at the apical surfaces of outer hair cells (Fig. 2m), and subsequently diminishes (Fig. 2o). In addition, PTK7 was also localized to the adherens junctions, a distribution which did not seem to change during development (Fig. 2n, p).

The phenotypic similarities between PTK7 and known PCP mutants strongly suggest that PTK7 is a novel regulator of PCP in vertebrates. To test for genetic interactions between PTK7 and the known PCP pathway, we crossed PTK7 heterozygous animals to *Lp* heterozygous animals and examined neural tube closure in the progeny. Whereas both single heterozygotes showed very low frequencies of NTD in the form of spina bifida, 94% of the double heterozygotes exhibited spina bifida (Fig. 3). This interaction was specific, as PTK7 heterozygotes crossed to either wild-type littermates of the *Lp* animals or mice heterozygous for the *patched* mutation¹², did not produce any progeny with NTD (data not shown). This trans-heterozygous interaction provides evidence that PTK7 and *vangl2* function in a common genetic pathway to regulate PCP. Interestingly, the interaction does not seem as strong as that reported between *Lp* and *Crc* mice, where a subset of *Lp*+/; *Crc*+/ embryos displayed the more severe craniorachischisis¹⁷. In addition, unlike in the *Lp*+/; *Crc*+/ embryos, stereociliary bundles of IHCs and OHC2s were largely unaffected in the *Lp*+/; PTK7+/ embryos (data not shown)¹⁶. These observations suggest that the dosage of PTK7 is less limiting than that of *mScrib1* and/or that there may be different genetic modifiers in the two compound heterozygous strains. We also examined PTK7 immunostaining and did not observe any change in *Lp* mutant cochleae at E16.5 (data not shown).

The PCP pathway in frogs and fish seems to mediate noncanonical, β -catenin-independent Wnt signalling to regulate convergent extension movements during gastrulation and neurulation, a process in which a tissue narrows mediolaterally and lengthens rostro-caudally^{4,5}. PCP-mediated neural convergent extension has been shown to be required for neurulation¹⁸. The phenotypes of the mouse PTK7 mutation (shortened body axis, broader floor plate and wider somites with shorter spacing, Fig. 1j) are consistent with a defect in convergent extension. To further test the involvement of PTK7 in vertebrate PCP, we used *Xenopus* to directly assess the role of PTK7 in convergent extension. By whole-mount *in situ* hybridization, the *Xenopus* homologue of PTK7 (*Xptk7*) was expressed in the neuroectoderm from late gastrula through all neurula stages (Fig. 4a). We used morpholino oligonucleotides to *Xptk7* to 'knock-down' PTK7 function in *Xenopus* embryos. Injection of *Xptk7* morpholino oligonucleotides, but not control morpholino oligonucleotides, blocked neural tube closure (Fig. 4b, c), a defect that could be rescued by coinjection of a full-length *Xptk7* construct that lacked the morpholino target site (data not shown). Overexpression of an *Xptk7* construct encoding a version of PTK7 lacking its cytoplasmic domain (*Xptk7* Δ C), but not of a full-length construct, resulted in similar phenotypes (Fig. 4b, c, and data not shown). Together, these results provide evidence for a specific requirement of PTK7 for neural tube closure. Defects in neural convergent extension in morpholino oligonucleotide-injected embryos were also apparent as demonstrated by *in situ* hybridization to different neural markers (Fig. 4d). Not only were the neural folds broader and more widely separated, but labelled cells also appeared more dispersed along the medial-lateral axis (for example, compare the

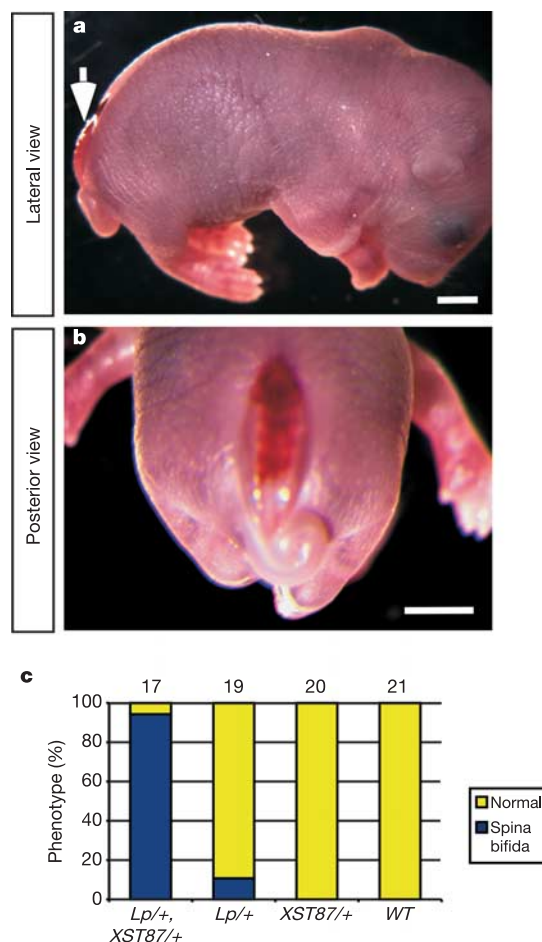


Figure 3 The PTK7 mutation genetically interacts with the *Lp* mutation. **a**, **b**, Lateral (**a**) and posterior (**b**) views of an E18.5 compound heterozygous embryo showing spina bifida (arrow in **a**). Scale bars, 2 mm. **c**, quantification of neural tube closure phenotypes in the progeny of the PTK7 heterozygotes (the gene trap allele is referred to as *XST87*) crossed to *Lp*+/ mice, the numbers of embryos for each of the four genotypes are indicated on top. WT, wild-type; normal, normal neural tube closure.

boxed areas in Fig. 4d, and magnified in Fig. 4e), indicative of a defect in mediolateral cell intercalations. We used time-lapse video microscopy to follow neural tube closure. An abnormally broad neural plate was already apparent at the onset of neurulation at Stage 14 (and at Stage 16 with the nascent neural folds) accompanied by a defect in axis elongation (Fig. 4f). Although in severe cases the entire neural tube remained open (data not shown), many morpholino oligonucleotide-injected embryos eventually closed their neural tubes despite a delay, with a dorsally flexed, shortened trunk (Fig. 4f). Taken together, these results indicate, as predicted, that *Xptk7* regulates neural convergent extension and neural tube closure.

Our results have identified PTK7 as a novel regulator of PCP in vertebrates. Because our conclusions are based principally on loss-of-function data, PTK7 could be either a component of the PCP pathway *per se* or involved in a parallel pathway required for PCP. However, our findings that PTK7 is required in disparate processes involving the PCP pathway (convergent extension/neural tube closure and stereociliary bundle orientation) and that it shows heteroallelic interactions with *Lp* lead us to favour a direct role for PTK7 as a PCP pathway component; if instead it is in a parallel pathway, it must be one that operates in multiple cellular contexts. Like most regulators of the PCP pathway, the precise signalling role of PTK7 remains to be determined. The *Drosophila* homologue of

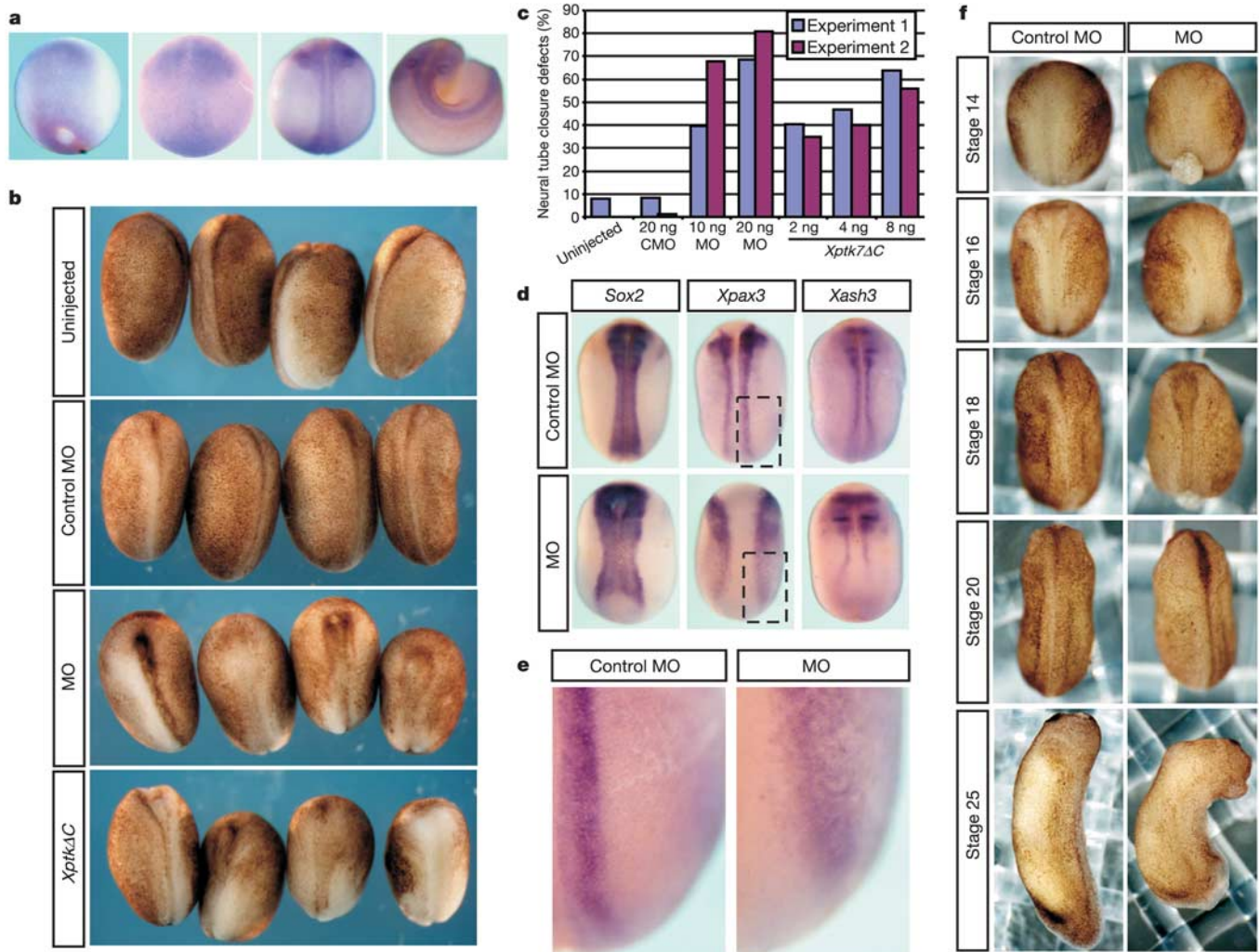


Figure 4 PTK7 regulates convergent extension and neural tube closure in *Xenopus*. Anterior is up. **a**, expression of *Xptk7* in neuroectoderm at stages 12.5, 15, 17 and 24, detected with antisense but not sense *Xptk7* riboprobe. **b**, injection of morpholino oligonucleotides targeting *Xptk7* (MO), or of *Xptk7ΔC* mRNA, but not control morpholino oligonucleotides, cause neural tube closure defects. **c**, quantification of defects. Two independent experiments ($n = 1,144$) are presented. CMO, control morpholino oligonucleotides. **d, e**, expression pattern of *Sox2* (pan-neural), *Xpax3* (lateral neural

plate) and *Xash3* (intermediate neural plate). Higher magnification (**e**) of boxed areas in **d**. In morpholino oligonucleotide-injected embryos, all markers showed shorter and wider-spaced expression domains. Furthermore, labelled cells are more dispersed along the medial-lateral axis, indicating defective neural convergent extension. **f**, time-lapse images of neural tube closure. Morpholino oligonucleotide-injected embryo showed greater distances between the neural folds, and a shorter body axis, as early as the onset of neurulation at Stage 14.

PTK7, *off-track* (*otk*), has been implicated as a co-receptor of plexins to mediate semaphorin signalling during axon guidance¹⁹. At present it is unclear whether *otk* is the functional orthologue of PTK7, as the *otk* mutation is lethal and possible PCP phenotypes in these fly mutants have not been characterized; conversely, the involvement of PTK7 in plexin signalling for axon guidance in vertebrates remains to be determined. The domain structure of PTK7 suggests that it could regulate PCP by adhesion (through the extracellular immunoglobulin domains) and/or signalling (through the cytoplasmic region with kinase homology) (Fig. 1b). Consistent with adhesion, mouse PTK7 can form multimers when over-expressed in cultured cells (data not shown), and homotypic interactions have been observed for OTK²⁰. As regards signalling, PTK7 is an atypical receptor tyrosine kinase (RTK) in that the DFG triplet that is essential for kinase activity is not conserved, so PTK7 lacks endogenous kinase activity²¹; however, the overall kinase homology is nonetheless conserved from hydra to humans²², suggesting that the cytoplasmic domain is functionally important. Expression of a mutant *Xptk7* that lacked the cytoplasmic domain caused NTD, demonstrating the importance of this domain; how-

ever, because both gain- and loss-of-function manipulations of PCP signalling components disrupt convergent extension⁵, we could not distinguish whether this mutation was activating or inactivating. Interestingly, immunoprecipitates of OTK appeared to contain tyrosine kinase activity, suggesting that OTK (and perhaps other PTK7 family members) might recruit other tyrosine kinase(s). The identification of PTK7 as a novel regulator of PCP provides an important entry point into the PCP and noncanonical Wnt signalling pathways in mammals. □

Methods

PTK7 and *Lp* mice

A mouse embryonic stem cell line carrying an insertion in the *PTK7* gene (XST87) was isolated and identified as described¹⁰. Mice carrying the *PTK7* insertion were generated by standard blastocyst injection and chimaera breeding. Genotyping of *PTK7* was performed by genomic PCR using two pairs of primers: for the wild-type allele, 5'-ACGTCAGTGG AGAGGAAGTCCGACG-3' and 5'-AGGACCACAGGACCTGCTTCGATC-3'; for the gene trap insertion allele, 5'-TGCACATGCTTTACGTGTG-3' and 5'-TGCCGCGTGTG GTGTTGCAC-3'. *PTK7* mice crossed to *Lp* mice had been backcrossed to C57BL/6 for five generations. *Lp* mice of the LPT/Le stock were obtained from the Jackson Laboratories. The following primers were used for genotyping of the *Lp* allele: 5'-AAACAGTGGACC TTGTTGAA-3' and 5'-AAGAGGAAGTAGGACTGGCAG-3'. For northern blot analysis,

a DNA fragment containing the entire open reading frame (ORF) of mouse PTK7 was used as a probe.

Histology, immunostaining and scanning electron microscopy

Histology and X-Gal staining was performed as described⁷. For immunostaining of the sensory hair cells, cochleae were dissected out of the temporal bones and the anlage of Reissner's membrane removed to expose the organ of Corti. Cochleae were then stained, flat-mounted and imaged as described¹⁶. Rabbit anti-PTK7 antibodies were raised against a PTK7 ecto-domain-Fc fusion protein (Zymed). Rhodamine-conjugated phalloidin was from Molecular Probes. Other antibodies were: anti-acetylated tubulin and anti-E-cadherin (Sigma); anti-PAX7, anti-LIM1/2, anti-PAX6, anti-Islet-1, anti-Nkx2.2 and anti-HNF3 β (Developmental Studies Hybridoma Bank); anti-Math1 (gift from J. Johnson), anti-LH2 (gift from T. Jessell) and Cy3- and Cy2-conjugated secondary antibodies (Jackson Immuno Laboratories). Images were collected using a Bio-Rad MRC 1024 laser scanning confocal microscope and LaserSharp image collection software. Orientation of individual bundle was determined by the vertex relative to the neural–abneural axis. 0° was defined as perpendicular to the row of pillar cells. Quantification of bundle orientation in phalloidin-stained E18.5 cochleae was performed essentially as described¹⁶. 315 control and 951 mutant hair cells sampled from the basal and middle turns of the cochleae of six embryos, respectively, were scored. Bundle orientation in the apical turns was not quantified due to an extra row of OHCs in the mutant. Scanning electron microscopy images of embryos and cochleae were taken at 20 kV on a Philips 525 scanning electron microscope upgraded with Semiscaps digital image acquisition. Samples were prepared using standard protocols.

Xenopus assays

RNA synthesis, *in situ* hybridization, culturing and injections were performed as described²³. *Xptk7* coding sequences were cloned by RT–PCR from mixed-staged *Xenopus* total RNA. *Xptk7* morpholino oligonucleotides complementary to sequences surrounding the translational start site and control morpholino oligonucleotides were synthesized by Gene Tools (LLC): (1) 5′-CGCAGAGAGACAATCGGCCCAT-3′; (2) 5′-TGCATCGCGGCCTCTCCCTCAC-3′; (3) 5′-TTCTGCGCCCGGATCCTCTCAGTGC-3′; and control morpholino oligonucleotide, 5′-CGAGAGAGCAATCGCCCCGAT-3′ (mismatched bases in lower case). Each *Xptk7* morpholino oligonucleotide alone caused NTD, with a higher potency when injected in combination. Neural tube closure of de-ventilated embryos was scored at stage 19–20. Embryos with completely closed neural tubes were scored as normal; those otherwise were scored as NTD. Embryos with severe gastrulation defects were not scored. Time-lapse microscopy analysis was carried out as described¹⁸.

Received 12 April; accepted 24 May 2004; doi:10.1038/nature02677.

- Adler, P. N. Planar signaling and morphogenesis in *Drosophila*. *Dev. Cell* **2**, 525–535 (2002).
- Strutt, D. Frizzled signalling and cell polarisation in *Drosophila* and vertebrates. *Development* **130**, 4501–4513 (2003).
- Mlodzik, M. Planar cell polarization: do the same mechanisms regulate *Drosophila* tissue polarity and vertebrate gastrulation? *Trends Genet.* **18**, 564–571 (2002).
- Keller, R. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* **298**, 1950–1954 (2002).
- Wallingford, J. B., Fraser, S. E. & Harland, R. M. Convergent extension: the molecular control of polarized cell movement during embryonic development. *Dev. Cell* **2**, 695–706 (2002).
- Veeman, M. T., Axelrod, J. D. & Moon, R. T. A second canon. Functions and mechanisms of beta-catenin-independent Wnt signaling. *Dev. Cell* **5**, 367–377 (2003).
- Leighton, P. A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
- Park, S. K., Lee, H. S. & Lee, S. T. Characterization of the human full-length PTK7 cDNA encoding a receptor protein tyrosine kinase-like molecule closely related to chick KLG. *J. Biochem. (Tokyo)* **119**, 235–239 (1996).
- Mossie, K. *et al.* Colon carcinoma kinase-4 defines a new subclass of the receptor tyrosine kinase family. *Oncogene* **11**, 2179–2184 (1995).
- Mitchell, K. J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nature Genet.* **28**, 241–249 (2001).
- Copp, A. J., Greene, N. D. & Murdoch, J. N. The genetic basis of mammalian neurulation. *Nature Rev. Genet.* **4**, 784–793 (2003).
- Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
- Hamblet, N. S. *et al.* Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. *Development* **129**, 5827–5838 (2002).
- Curtin, J. A. *et al.* Mutation of *Celsr1* disrupts planar polarity of inner ear hair cells and causes severe neural tube defects in the mouse. *Curr. Biol.* **13**, 1129–1133 (2003).
- Bilder, D., Schober, M. & Perrimon, N. Integrated activity of PDZ protein complexes regulates epithelial polarity. *Nature Cell Biol.* **5**, 53–58 (2003).
- Montcouquiol, M. *et al.* Identification of *Vangl2* and *Scrb1* as planar polarity genes in mammals. *Nature* **423**, 173–177 (2003).
- Murdoch, J. N. *et al.* Circletail, a new mouse mutant with severe neural tube defects: chromosomal localization and interaction with the loop-tail mutation. *Genomics* **78**, 55–63 (2001).
- Wallingford, J. B. & Harland, R. M. Neural tube closure requires Dishevelled-dependent convergent extension of the midline. *Development* **129**, 5815–5825 (2002).
- Winberg, M. L. *et al.* The transmembrane protein Off-track associates with Plexins and functions downstream of Semaphorin signaling during axon guidance. *Neuron* **32**, 53–62 (2001).
- Pulido, D., Campuzano, S., Koda, T., Modolell, J. & Barbacid, M. *Dtrk*, a *Drosophila* gene related to the *trk* family of neurotrophin receptors, encodes a novel class of neural cell adhesion molecule. *EMBO J.* **11**, 391–404 (1992).
- Kroher, M., Miller, M. A. & Steele, R. E. Deceiving appearances: signaling by “dead” and “fractured” receptor protein-tyrosine kinases. *Bioessays* **23**, 69–76 (2001).
- Miller, M. A. & Steele, R. E. Lemon encodes an unusual receptor protein-tyrosine kinase expressed

during gametogenesis in Hydra. *Dev. Biol.* **224**, 286–298 (2000).

- Borchers, A. G. *et al.* The E3 ubiquitin ligase GREUL1 anteriorizes ectoderm during *Xenopus* development. *Dev. Biol.* **251**, 395–408 (2002).
- Murdoch, J. N. *et al.* Disruption of scribble (*Scrb1*) causes severe neural tube defects in the circletail mouse. *Hum. Mol. Genet.* **12**, 87–98 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Perrino and J. Mulholland at the Stanford Cell Sciences Imaging Facility for assistance with scanning electron microscopy analysis; J. Zhong for technical support; R. M. Harland, J. E. Johnson, T. M. Jessell and M. P. Scott for reagents; D. Brown for helpful discussions; and C. I. Bargmann, S. K. McConnell, J. D. Axelrod, M. A. Simon, T. Venkatesh, U. Grieshammer, L. V. Goodrich and members of the Tessier-Lavigne laboratory for helpful comments on the manuscript. Funding for this project was provided by grants to M.T.L. from the NIMH, and to J.C.B. from the NIH. X.L. was supported by a fellowship from the Damon Runyon Cancer Research Foundation, and A.G.M.B. by the Stanford University Dean's Fellowship. M.T.L. was an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.T.L. (marctl@gene.com).

Nitrification by plants that also fix nitrogen

Charles R. Hipkin*, Deborah J. Simpson*†, Stephen J. Wainwright & Mansour A. Salem‡

School of Biological Sciences, University of Wales Swansea, Institute of Environmental Sustainability, Swansea SA2 8PP, UK

† Present address: Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3TL, UK

‡ Present address: Faculty of Engineering and Technology, Sebha University, Sebha, Libya

* These authors contributed equally to this work.

Nitrification is a key stage in the nitrogen cycle; it enables the transformation of nitrogen into an oxidized, inorganic state^{1,2}. The availability of nitrates produced by this process often limits primary productivity and is an important determinant in plant community ecology and biodiversity^{3–6}. Chemoautotrophic prokaryotes are recognized as the main facilitators of this process⁷, although heterotrophic nitrification by fungi may be significant under certain conditions⁸. However, there has been neither biochemical nor ecological evidence to support nitrification by photoautotrophic plants. Here we show how certain legumes that accumulate the toxin, 3-nitropropionic acid, generate oxidized inorganic nitrogen in their shoots, which is returned to the soil in their litter. In nitrogen-fixing populations this ‘new’ nitrate and nitrite can be derived from the assimilation of nitrogen gas. Normally, the transformation of elemental nitrogen from the atmosphere into a fixed oxidized form (as nitrate) is represented in the nitrogen cycle as a multiphase process involving several different organisms. We show how this can occur in a single photoautotrophic organism, representing a previously undescribed feature of this biogeochemical cycle.

A number of leguminous plants (Fabaceae) are known to accumulate free and esterified forms of nitroaliphatic compounds, such as 3-nitropropionic acid (3NPA) and 3-nitropropanol^{9–14}. Only a small proportion of this large family of plants has been surveyed so far, but these compounds have been found in over 500 species representing the following seven genera in the sub-family Papilionoideae: *Astragalus*, *Coronilla*, *Hippocrepis*, *Indigofera*, *Lotus*, *Securigera* and *Scorpiurus*. In some species, concentrations of 3NPA can exceed 100 μ mol per g fresh weight¹⁵, representing a significant proportion (>7%) of the soluble fixed nitrogen in shoots. Because 3NPA is a potent inhibitor of succinate dehydrogenase¹⁶, a key

enzyme in respiratory and intermediate metabolism, it is a notable toxin that probably plays a major part in plant defence against herbivory. In horseshoe vetch (*Hippocrepis comosa*) it accumulates in all parts of the shoot^{14,15,17} but not in roots or suspension cultures. It is metabolized in an enzymic oxidation process catalysed by 3-nitropropionic acid oxidase (3NPAOx) (ref. 17) which prevents auto-inhibition, a potential problem when 3NPA leaks into vascular tissue and sensitive compartments of the cell. This happens when leaves are damaged or during senescence, before leaf shedding. In this reaction two equivalents of nitrate and one of nitrite are produced for every molecule of 3NPA oxidized (see Fig. 1 caption). In nodulated leguminous plants, nitrogen compounds such as 3NPA may be derived from the assimilation of ammonium generated by the biological fixation of atmospheric nitrogen. Therefore, this reaction makes the transformation of nitrogen gas into oxidized inorganic nitrogen, by means of nitrogen fixation and nitrification, possible in a single organism. This process is not described as part of the terrestrial nitrogen cycle, but if it occurred in all leguminous species that accumulate 3NPA it would be a widespread phenomenon. This would only be feasible if 3NPAOx activity is widespread in leguminous species that accumulate 3NPA and if oxidized inorganic nitrogen generated by this process *in vivo* is returned to the nitrogen cycle in the soil. Here we describe the results of experiments designed to test these possibilities.

A survey of over 200 species of plants, including 82 legumes, revealed a striking correspondence between 3NPA accumulation and 3NPAOx activity. Table 1 shows the results obtained from 44 leguminous species. Nitrogen-fixing species in the genera *Astraga-*

lus, *Coronilla*, *Hippocrepis*, *Lotus*, *Scorpiurus* and *Securigera*, which accumulated 3NPA, also contained 3NPAOx activity in leaf extracts. In contrast, 3NPAOx activity was not found in species that did not accumulate 3NPA. This included legumes such as white clover (*Trifolium repens*) and all non-leguminous species tested (list of species elsewhere¹⁴).

These results indicate the widespread occurrence of a nitrification mechanism in legumes. However, this has no significance in terms of terrestrial nitrification unless oxidized inorganic nitrogen is generated *in vivo* by this process in natural populations and then returned to the soil. To investigate this we measured the production of 3NPA, nitrate and nitrite in horseshoe vetch plants grown in water culture and sampled from natural populations (see Methods). Initially, nitrate, nitrite and 3NPA were measured in healthy, mature leaves collected from non-nodulated plants that had only ever been grown in cultures containing nitrogen in the reduced combined forms of urea or ammonium (Table 2). Of the nitrogen partitioned between 3NPA, nitrite and nitrate (collectively described here as nitro-nitrogen), most was present as 3NPA. However, up to 20% of nitro-nitrogen was present as oxidized inorganic nitrogen (nitrite and nitrate) in these plants. It is notable that this oxidized inorganic nitrogen is new (that is, not recycled) nitrate and nitrite that has been derived from the assimilation of reduced combined nitrogen. We have also shown that new nitro-nitrogen is produced by axenic cultures of horseshoe vetch (Supplementary Information).

Nitro-compounds were also measured in leaves from nodulated plants that were fixing nitrogen in water culture. These plants contained the highest levels of 3NPA that we have recorded (814 $\mu\text{mol per g dry weight}$) whereas oxidized inorganic nitrogen (90 $\mu\text{mol per g dry weight}$) represented 11% of the total nitro-nitrogen. We also found that leaves from nodulated plants in grassland populations on the cliffs of the south Gower peninsular at Limeslade and Westcliff contained high concentrations of 3NPA (222–473 $\mu\text{mol per g dry weight}$) with up to 27% of the total nitro-nitrogen partitioned in the oxidized inorganic form.

One might expect the nitrate and nitrite generated during the oxidation of 3NPA in senescent leaves to be re-assimilated and translocated before litter production, in order to conserve fixed nitrogen. Thus, total amounts of nitro-nitrogen were lower in litter compared with in fresh leaves (Table 2). However, litter always contained large residual amounts of nitro-nitrogen under all growth conditions. Litter from natural populations contained 12.6–68 $\mu\text{mol nitro-nitrogen per g dry weight}$, of which up to 67% was present as oxidized inorganic nitrogen, predominantly in the form of nitrate (Table 2; Supplementary Table 1). In three consecutive summers at Limeslade and two consecutive summers at Westcliff, the oxidized inorganic nitrogen content of horseshoe vetch litter varied from 8.5 to 29.1 $\mu\text{mol per g dry weight}$. Litter capture

Table 1 The occurrence of 3NPA and 3NPAOx in legumes

Species	3NPA	3NPAOx activity
<i>Anthyllis vulneraria</i>	0	0
<i>Astragalus alpinus</i>	0	0
<i>A. cicer</i>	0	0
<i>A. danicus</i>	0	0
<i>A. falcatus</i>	69.8(±2.6)	15.5(±1.2)
<i>A. lusitanicus</i>	0	0
<i>A. mexicanus</i>	0	0
<i>Coronilla coronata</i>	0	0
<i>C. juncea</i>	0	0
<i>C. minima</i>	0	0
<i>C. scorpioides</i>	0	0
<i>C. vaginalis</i>	0	0
<i>C. valentina</i>	0	0
<i>C. valentina glauca</i>	0	0
<i>C. viminalis</i>	38.5(±1.8)	561(±109)
<i>Hippocrepis balearica</i>	24.7 (±4.2)	5522 (±144)
<i>H. comosa</i>	142(±7.5)	173(±8.2)
<i>H. emerus emerus</i>	60.8(±3.3)	90.8(±10.5)
<i>H. unisiliquosa</i>	61.5(±5.7)	1408(±21)
<i>Lotus alpinus</i>	0	0
<i>L. angustissimus</i>	6.9(±0.6)	10.7(±0.5)
<i>L. corniculatus</i>	0	0
<i>L. creticus</i>	0	0
<i>L. edulis</i>	0	0
<i>L. pedunculatus</i>	13.8(±2.2)	11.0(±1.0)
<i>L. tenuis</i>	0	0
<i>L. tetragonolobus</i>	0	0
<i>L. uliginosus</i>	63.3(±4.0)	118(±2.2)
<i>Onobrychis arenaria</i>	0	0
<i>O. montana</i>	0	0
<i>Oxytropis campestris</i>	0	0
<i>O. gaudinii</i>	0	0
<i>O. jacquinii</i>	0	0
<i>Scorpiurus muricatus</i>	9.4(±0.6)	518(±27.2)
<i>S. vermiculatus</i>	18.0(±0.9)	36.5(±1.8)
<i>Securigera orientalis</i>	22.3(±0.8)	263(±63.3)
<i>S. parviflora</i>	47.4 (±12.4)	929 (±2.3)
<i>S. securidaca</i>	41.9(±1.7)	812(±15.8)
<i>S. varia</i>	85.9(±6.3)	1517(±91.0)
<i>Tetragonolobus purpureus</i>	0	0
<i>Thermopsis caroliniana</i>	0	0
<i>Trifolium repens</i>	0	0
<i>Trigonella foenum-graecum</i>	0	0
<i>Vicia dalmatica</i>	0	0

3NPA is given as $\mu\text{mol g}^{-1}$ fresh weight. 3NPAOx activity is expressed as $\text{nmol NO}_2 \text{ min}^{-1}$ per mg protein ($n = 3$, $\pm$ s.e.m.).

Table 2 Concentrations of 3NPA, nitrite and nitrate in horseshoe vetch

	3NPA	NO ₂	NO ₃
(a) In mature leaves			
NH ₄ water culture	265.1(±7.1)	30.3(±1.0)	13.9(±4.0)
Urea water culture	214(±6.0)	43.2(±1.5)	11.4(±4.0)
N ₂ water culture	814(±14.3)	38.0(±0.7)	51.9(±4.2)
Limeslade 1995	221.8(±5.1)	54.4(±0.56)	29.1(±3.5)
Limeslade 1996	292.8(±15.3)	29.8(±1.2)	29.2(±2.2)
Limeslade 1997	487.1(±6.4)	60.4(±0.5)	23.0(±2.0)
Westcliff 1996	472.8(±25.7)	49.7(±1.5)	6.8(±1.2)
Westcliff 1997	401.3(±12.9)	40.9(±2.8)	19.2(±1.1)
(b) In leaf litter			
NH ₄ water culture	70.3(±2.3)	5.1(±1.3)	24.5(±1.3)
Urea water culture	30.9(±1.3)	2.2(±0.1)	14.9(±0.1)
Limeslade 1995	4.1(±0.1)	1.4(±0.1)	7.1(±0.1)
Limeslade 1996	34.2(±1.8)	1.4(±0.4)	14.9(±0.8)
Limeslade 1997	38.9(±2.8)	2.0(±0.12)	27.1(±2.5)
Westcliff 1996	36.8(±2.6)	3.2(±0.2)	5.6(±0.7)
Westcliff 1997	14.5(±4.8)	4.4(±0.42)	12.3(±1.0)

Concentrations of 3NPA, NO₂ and NO₃ are given as $\mu\text{mol per g dry weight}$. ($n = 3$, $\pm$ s.e.m.)

Table 3 **Oxidized nitrogen in the litter from natural populations of horseshoe vetch**

Nitro-compound	Limeslade	Westcliff
Total nitro-nitrogen	44.9	17.0
Nitrate-nitrogen	15.3	5.5
Nitrite-nitrogen	1.3	1.7
Litter	4.4	2.4
$\delta^{15}\text{N}_t$ (horseshoe vetch)	$-1.51(\pm 0.85)$	$-1.56(\pm 0.54)$
$\delta^{15}\text{N}_s$ (reference plants)	$+2.17(\pm 0.62)$	$+0.86(\pm 0.24)$
% Fixed nitrogen derived from the assimilation of atmospheric nitrogen	88.2	84.6

Populations of horseshoe vetch were collected in the quarterly period between June and September 1997. Concentrations of nitro-compounds are given as mg N m^{-2} . Total nitro-nitrogen includes 3NPA + NO_3^- + NO_2^- . Litter is given as g m^{-2} . For $\delta^{15}\text{N}$ analysis, Limeslade horseshoe vetch ($n = 9$), reference plants ($n = 5$); Westcliff horseshoe vetch ($n = 10$), reference plants ($n = 8$). The differences between horseshoe vetch and reference plants were significant in both communities (Limeslade $P = 0.004$; Westcliff $P = 0.0009$). $\delta^{15}\text{N}$ is expressed as ‰.

experiments were set up to measure the amount of oxidized inorganic nitrogen being returned to the soil by means of horseshoe vetch litter at each of our sites (Table 3). Horseshoe vetch litter is easily identified for collection, although its mobility in these exposed sites made it difficult to ensure the capture of all litter. In view of this our figures for returned nitrogen must be underestimates. Even so, the results show that an average of 12 mg N (as NO_3^- + NO_2^-) was returned in leaf litter per square metre of grassland in a 13-week (quarterly) period between June and September. Because horseshoe vetch litter is generated throughout the year, this indicates an annual nitrogen deposition of at least 48 mg m^{-2} .

Whereas nitrogen fixation facilitates the entry of nitrogen into the nitrogen cycle from a huge atmospheric reservoir of nitrogen gas, nitrification is part of a complex soil mineralization process that generates new nitrate. These processes are normally regarded as discrete phases in the nitrogen cycle involving specialist organisms, separated by an obligatory phase of ammonification. The possibility that nitrogen can be fixed, assimilated into organic form and transformed into new nitrate in the nitrogen cycle in a single organism has never been explored. Here we provide evidence for a non-microbial nitrification in which new nitrate can be generated from atmospheric nitrogen fixed by symbiotic autotrophic plants, and enter the nitrogen cycle in litter. Because the accumulation and oxidation of 3NPA only occurs in the photoautotrophic tissues of the shoot¹⁷, we refer to this process as photoautotrophic nitrification. Nitrate encapsulated in litter would be released slowly as litter is leached and decomposed. The process is summarized in Fig. 1. Because nitrate and ammonium in soil may also be available to legumes, the assimilation of these nitrogen sources into 3NPA is also indicated. However, although photoautotrophic nitrification occurs under all these conditions, only nitrate derived from the assimilation of ammonium or atmospheric nitrogen, qualifies as new nitrate in this model. Measurements of the natural abundance of ^{15}N in horseshoe vetch ($\delta^{15}\text{N}_t$, the $\delta^{15}\text{N}$ value of the total nitrogen in the nitrogen-fixing plant grown under conditions in which atmospheric nitrogen and nitrogen from other sources are available) indicate that at least 85% of the fixed nitrogen in the Gower populations is derived from the assimilation of atmospheric nitrogen (Table 3). Measurements of nitrogen isotope abundances in other temperate grassland and forest ecosystems also indicate the predominance of nitrogen fixation in legume populations¹⁸. Ammonium may be an important source of nitrogen, particularly for species of wet, acidic grasslands (for example, *Lotus pedunculatus*) or acidic prairies (*Astragalus* spp.). In these ecosystems new nitrate, generated by photoautotrophic nitrification, could originate from nitrogen derived from the atmosphere and/or ammonium derived from the soil. At present the metabolic pathway of 3NPA synthesis is not understood. However, roots of species like horseshoe vetch export a large proportion of their organic nitrogen as asparagine (Supplementary Table 2), which would generate aspartic acid, a suggested precursor for 3NPA¹⁹.

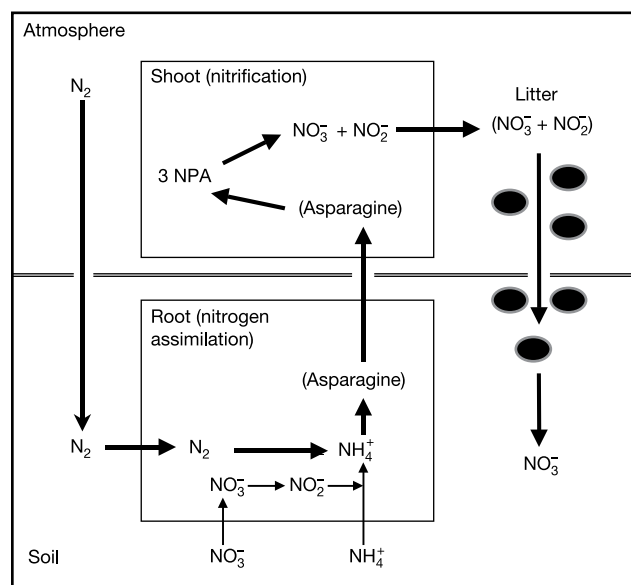


Figure 1 Nitrogen assimilation and photoautotrophic nitrification in legumes. Nitrogen (N_2) from the atmosphere, or combined inorganic nitrogen from the soil (NH_4^+ or NO_3^-), may be assimilated and converted to oxidized inorganic nitrogen (predominantly NO_3^-) in shoot litter. Nitrogen is fixed into an organic form (for example, asparagine) in the root and transported to the shoot. During senescence 3NPA is oxidized to NO_3^- and NO_2^- in deciduous shoot tissues and returned to the soil in litter. The nitrification reaction catalysed by 3NPAox is as follows: $3\text{NO}_2(\text{CH}_2)_2\text{CO}_2\text{H} + 3\text{O}_2 \rightarrow 3\text{CHOCH}_2\text{CO}_2\text{H} + \text{NO}_2^- + 2\text{NO}_3^- + (\text{H}_2\text{O})_x$

Whereas the contribution from photoautotrophic nitrification to the global nitrogen cycle can only be a small fraction of that contributed by chemoautotrophic nitrification, our results indicate that a biochemical mechanism for nitrification in legumes is widespread. On the basis of field experiments, our estimates suggest that nitrate input into horseshoe vetch litter is considerably less than the rates of microbial mineralization that occur in some non-agricultural soils^{20,21}, which sometimes exceeds $1 \text{ g N m}^{-2} \text{ yr}^{-1}$. However, field measurements reveal a great variation in natural rates of microbial chemoautotrophic nitrification^{20,22}, which are frequently lower than those reported here for photoautotrophic nitrification. Moreover, nitrogen-fixing legumes often occur in nitrogen-deficient ecosystems or ecosystems where chemoautotrophic nitrification is compromised. In these ecosystems the occurrence of 3NPA-containing legumes provides an alternative nitrification process, uncoupled from the microbial community and not limited by the size of the ammonium pool in the soil. □

Methods

Plant material and measurements of ^{15}N abundance

Natural populations of *H. comosa* L. were studied at two limestone grassland sites on south Gower in south Wales; Limeslade (UK ordnance survey grid reference number SS628872) and Westcliff (SS552872). The litter that accumulated in a 13-week period between June and September 1997 was collected *in situ* from quadrats (625 cm^2) that had been staked out at each site (six quadrats at Westcliff and five at Limeslade).

Natural ^{15}N abundance was measured in leaves of horseshoe vetch and reference plants (*Centaurea scabiosa*, *Erodium cicutarium* and *Sanguisorba minor*) using a PDZ Europa 2020 continuous flow stable isotope mass spectrometer. Reference plant selection, sample preparation protocols and the calculation of the fixed nitrogen derived from the assimilation of atmospheric nitrogen ($\delta^{15}\text{N}_a = -2.0 \text{ ‰}$) were as described in ref. 23.

Non-nodulated plants were maintained and grown in a water culture solution²⁴ supplied with ammonium chloride (1.5 mM) or urea (0.75 mM) as the sole source of nitrogen. Plants initially grown with ammonium were washed thoroughly in nitrogen-free solution and nodulated after infection with *Rhizobium* in a nitrogen-free perlite-based medium. Nodules collected from *H. comosa* roots in natural populations were washed in nitrogen-free medium, crushed, and used as a source of the bacteria. While in nitrogen-free culture, plants did not grow until they had become nodulated and then grew rapidly.

To ensure that new shoots were receiving reduced nitrogen resulting from nitrogen fixation, plants were cut back to the crown to remove shoot biomass derived from previous ammonium assimilation. This cycle was repeated and subsequent new growth was allowed to accrue over a period of 12 weeks.

Plant material used in the survey of 3NPA accumulation and 3NPAOx was collected directly in the field or from plants grown from seed obtained from botanic gardens in Europe and North America (see Acknowledgements).

Enzyme assays and analysis of nitro-compounds

3NPA in leaves was measured by the standard method²⁵. After cooling, nitrite was measured in hot water (80 °C) extracts of leaves using a modified Griess–Ilosvay assay²⁶. Nitrate was measured in these extracts after reduction to nitrite in a cadmium/copper column²⁷.

The activity of 3NPAOx was measured in partially purified leaf extracts. Two grams of fresh leaves were ground in 10 ml of 50 mM MES buffer at pH 5.0. After centrifugation at 10,000 g for 30 min at 4 °C, a 40–60% ammonium sulphate fraction was taken and used as the source of enzyme. Enzyme activity was assayed by nitrite production and oxygen consumption¹⁷.

Received 13 November 2003; accepted 4 May 2004; doi:10.1038/nature02635.

- Focht, D. D. & Verstraete, W. Biochemical ecology of nitrification and denitrification. *Adv. Microb. Ecol.* **1**, 135–214 (1977).
- Sprent, J. I. *The Ecology of the Nitrogen Cycle* (Cambridge Univ. Press, Cambridge, 1987).
- Chapin, F. S. The mineral nutrition of wild plants. *Annu. Rev. Ecol. Syst.* **11**, 233–260 (1980).
- Berendse, F. Interspecific competition and niche differentiation between *Plantago lanceolata* and *Anthoxanthum odoratum* in a natural hayfield. *J. Ecol.* **78**, 413–427 (1983).
- Verhoeven, J. T. A., Koerselman, W. & Meuleman, A. F. M. Nitrogen or phosphorus-limited growth in herbaceous wet vegetation: relations with atmospheric inputs and management regimes. *Trends Ecol. Evol.* **11**, 494–497 (1996).
- Wedin, D. A. & Tilman, D. Influence of nitrogen loading and species composition on carbon balance of grasslands. *Science* **274**, 1720–1721 (1996).
- Macdonald, R. M. in *Nitrification* (ed. Prosser, J. I.) 1–16 (IRL, Oxford, 1986).
- Killham, K. in *Nitrification* (ed. Prosser, J. I.) 117–126 (IRL, Oxford, 1986).
- Gustine, D. L. Aliphatic nitrocompounds in Crown Vetch: A review. *Crop Sci.* **19**, 197–203 (1979).
- Williams, M. C. & Barneby, R. C. The occurrence of nitro-toxins in North American *Astragalus* (Fabaceae). *Brittonia* **29**, 310–326 (1977).
- Williams, M. C. & Barneby, R. C. The occurrence of nitro-toxins in old world and South American *Astragalus* (Fabaceae). *Brittonia* **29**, 327–331 (1977).
- Williams, M. C. Nitrocompounds in *Indigofera* species. *Agron. J.* **73**, 434–436 (1981).
- Williams, M. C. Toxic nitro compounds in *Lotus*. *Agron. J.* **75**, 520–522 (1983).
- Salem, M. A., Williams, M., Wainwright, S. J. & Hipkin, C. R. Nitroaliphatic compounds in *Hippocrepis comosa* and other legumes in the European flora. *Phytochemistry* **40**, 89–91 (1995).
- Simpson, D. J., Wainwright, S. J. & Hipkin, C. R. Presence of 3-nitropropionic acid in widely distributed pasture legumes in Britain. *Vet. Rec.* **145**, 169–171 (1999).
- Alston, T. A., Porter, D. J. T. & Bright, H. J. Enzyme inhibition by nitro and nitroso compounds. *Acc. Chem. Res.* **16**, 418–424 (1983).
- Hipkin, C. R., Salem, M. A., Simpson, D. & Wainwright, S. J. 3-nitropropionic acid oxidase from horseshoe vetch (*Hippocrepis comosa*): a novel plant enzyme. *Biochem. J.* **340**, 491–495 (1999).
- Dijkstra, P. et al. Nitrogen stable isotope composition of leaves and roots of plants growing in a forest and a meadow. *Isotopes Environ. Health Stud.* **39**, 29–39 (2003).
- Baxter, R. L. et al. Fungal biosynthesis of 3-nitropropionic acid. *J. Chem. Soc. Perkin Trans. 1*, 2495–2501 (1992).
- Rehder, H. & Schäfer, A. Nutrient turnover studies in alpine ecosystems. IV. Communities of the Central Alps and comparative survey. *Oecologia* **34**, 309–327 (1978).
- Pastor, J., Stillwell, M. A. & Tilman, D. Nitrogen mineralization and nitrification in four Minnesota old fields. *Oecologia* **71**, 481–485 (1987).
- Morecroft, M. D., Marrs, R. M. & Woodward, F. I. Altitudinal and seasonal trends in soil nitrogen mineralization rate in the Scottish Highlands. *J. Ecol.* **80**, 49–56 (1992).
- Shearer, G. & Kohl, D. H. N₂-fixation in field settings: Estimations based on natural ¹⁵N abundance. *Aust. J. Plant Physiol.* **13**, 699–756 (1986).
- Arnon, D. I. & Hoagland, D. R. Crop production in artificial culture solutions and in soils, with special reference to factors influencing yield and absorption of inorganic nutrients. *Soil Sci.* **50**, 463–483 (1940).
- Matsumoto, H., Unrau, A. M., Hylin, J. W. & Temple, B. Spectrophotometric determination of 3-nitropropionic acid in biological extracts. *Anal. Chem.* **33**, 1442–1444 (1961).
- Everest, S. A., Hipkin, C. R. & Syrett, P. J. The effect of phosphate and flavin adenine dinucleotide on nitrate reductase activity of some unicellular marine algae. *J. Exp. Mar. Biol. Ecol.* **76**, 263–275 (1984).
- Parsons, T. R., Maita, Y. & Lalli, C. M. *A Manual of Chemical and Biological Methods for Seawater Analysis* (Pergamon, Oxford, 1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Williams, D. Knight, R. Jones and K. Flynn for their assistance. M.A.S. thanks the Libyan government for financial support. The following botanic gardens supplied seed: Akureyri, Iceland; Bayreuth, Germany; Bern, Switzerland; Bonn, Germany; Bordeaux, France; Bratislava, Slovakia; Caen, France; Cambridge, UK; Dresden, Germany; Firenze, Italy; Halle, Germany; Hamburg, Germany; Innsbruck, Austria; Jenna, Germany; L'Aquila, Italy; Latte, Italy; Leicester, UK; Leipsig, Germany; Lisbon, Portugal; Oldenburg, Germany; Oxford, UK; Padova, Italy; Rouen, France; Sienna, Italy; Strasbourg, France; and Urbino, Italy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.H. (c.r.hipkin@swansea.ac.uk).

Context-dependent contributions of backbone hydrogen bonding to β -sheet folding energetics

Songpon Deechongkit¹, Houbi Nguyen², Evan T. Powers¹, Philip E. Dawson^{1,3}, Martin Gruebele^{2,4} & Jeffery W. Kelly¹

¹Department of Chemistry, and ³Department of Cell Biology and The Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, BCC 506, La Jolla, California 92037, USA

²Center for Biophysics and Computational Biology, and ⁴Department of Chemistry, Physics, and Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana, Illinois 61801, USA

Backbone hydrogen bonds (H-bonds) are prominent features of protein structures; however, their role in protein folding remains controversial because they cannot be selectively perturbed by traditional methods of protein mutagenesis^{1–3}. Here we have assessed the contribution of backbone H-bonds to the folding kinetics and thermodynamics of the PIN WW domain, a small β -sheet protein⁴, by individually replacing its backbone amides with esters. Amide-to-ester mutations site-specifically perturb backbone H-bonds in two ways: a H-bond donor is eliminated by replacing an amide NH with an ester oxygen, and a H-bond acceptor is weakened by replacing an amide carbonyl with an ester carbonyl^{5–13}. We perturbed the 11 backbone H-bonds of the PIN WW domain by synthesizing 19 amide-to-ester mutants. Thermodynamic studies on these variants show that the protein is most destabilized when H-bonds that are enveloped by a hydrophobic cluster are perturbed. Kinetic studies indicate that native-like secondary structure forms in one of the protein's loops in the folding transition state, but the backbone is less ordered elsewhere in the sequence. Collectively, our results provide an unusually detailed picture of the folding of a β -sheet protein.

Several members of the WW domain family have been used as model systems to understand the kinetics and thermodynamics of β -sheet folding^{14–22}. The PIN WW domain is the 34-residue, independently folded, ligand-binding portion of the PIN1 protein, a two-domain phosphoserine/phosphothreonine–proline isomerase involved in cell signalling²³. The structure of the PIN WW domain, and the location of 10 of its 11 H-bonds relative to its hydrophobic core^{18,24}, are shown in Fig. 1. Traditional side chain mutagenesis studies, in which one amino acid residue is replaced by another, have shown that the PIN WW domain can tolerate side chain modification at nearly every position¹⁷. Thermodynamic and kinetic analyses of these mutants have identified two groups of residues: one important for thermodynamic stability, and another important for transition state (TS) energetics and thus the folding rate¹⁷. Several residues involved in hydrophobic core formation or electrostatic interactions in loop 2 contribute disproportionately to the stability of this protein and thus are in the first group¹⁷. Residues in loop 1, the formation of which dominates the TS, are in the second group¹⁷.

Amide-to-ester (A-to-E) mutants of PIN WW domain were prepared by using solid-phase synthesis based on a Boc/benzyl protecting group strategy, in which α -amino acid residues were individually replaced with their α -hydroxy acid analogues^{12,13}. Previous studies of backbone H-bonding using A-to-E mutants have been limited by the availability of appropriately protected α -hydroxy acid analogues of α -amino acids (side-chain-protected α -hydroxy acid analogues of arginine, aspartate, cysteine, glutamate, histidine, lysine, serine, threonine, tryptophan and tyrosine are not commercially available)^{5–13}. Therefore, to obtain all of the 19 A-to-E

mutants required to probe the 11 H-bonds in the PIN WW domain, we have chemically synthesized all of the necessary chiral, protected α -hydroxy acids²⁵. α -Hydroxy acid residues in protein sequences are denoted by lowercase Greek letters that correspond to the single-letter code of the corresponding α -amino acid residues (for example, 'α' represents lactic acid because 'A' represents alanine).

All of the A-to-E PIN WW domain variants were monomeric on the basis of sedimentation equilibrium analytical ultracentrifugation, and they all had a wild-type-like structure according to their far-ultraviolet circular dichroism, fluorescence emission and one-dimensional ¹H NMR spectra, and their ability to bind a proline-rich phosphoserine peptide with wild-type-like affinity (S.D., P.E.D. and J.W.K., unpublished results).

Thermodynamic data at 2 °C were derived from guanidine hydrochloride denaturation curves. For more destabilized A-to-E mutants, the osmolyte trimethylamine *N*-oxide was added to facilitate accurate measurements²⁶. The differences between the folding free energies of the wild-type PIN WW domain and the A-to-E mutants ($\Delta\Delta G_f$) are shown in Fig. 2. The overall average ($\pm$ s.d.) value of $\Delta\Delta G_f$ is -6.7 ± 5.8 kJ mol⁻¹ but, generally speaking, the values of $\Delta\Delta G_f$ fall into three ranges. When an A-to-E mutation merely weakens a buried or solvent exposed H-bond acceptor (by replacing an amide carbonyl with an ester carbonyl; K12κ, M15μ, R17ρ, V22ω, H27η, S32σ and W34ω), the average $\Delta\Delta G_f$ value is -2.5 ± 1 kJ mol⁻¹. When a solvent exposed donor is eliminated (by replacing it with an ester O; W11ω, E12ε, S16σ, S19σ, Y24ψ, N30ν, A31α), the average $\Delta\Delta G_f$ value is -4.8 ± 1.7 kJ mol⁻¹. Finally, when a buried donor is eliminated (R14ρ, Y23ψ, F25φ, N26ν and Q33θ), the average $\Delta\Delta G_f$ value is -15.2 ± 4.2 kJ mol⁻¹ (note that the status of the acceptor had a negligible effect on the $\Delta\Delta G_f$ values of A-to-E mutants in which a buried donor was removed).

A detailed analysis of the thermodynamic influence of the A-to-E mutations, including approaches towards extracting H-bond energies from the $\Delta\Delta G_f$ values by correcting them for electrostatic interactions and desolvation differences, will be published elsewhere (S.D., P.E.D. and J.W.K., unpublished results). Here, we note that the location of a backbone H-bond dictates the extent of protein destabilization when the H-bond donor is eliminated by an A-to-E

mutation. The five buried H-bonds shown in red in Fig. 1, for which elimination of the donors considerably destabilizes the protein, are in the hydrophobic core. By contrast, A-to-E mutations that perturb solvent-exposed backbone H-bonds near or in the loops and at the ends of the strands had much smaller effects on stability of the PIN WW domain, owing to the high dielectric constant of the micro-environment and the ability of water to compete with the intramolecular H-bond. This is consistent with results from A-to-E mutants of GCN4, an α -helical coiled-coil domain¹³.

Thermal denaturation-based thermodynamic and laser temperature-jump (T-jump) kinetic data were used together to discern the contribution of backbone H-bonding to TS energetics, as has been done for side chain mutants^{16,17,20}. The wild-type PIN WW domain folds very fast, owing to its small size and simple topology (the folding half-life is about 100 μs at 50 °C)¹⁷. The influence of an A-to-E mutation on TS energetics relative to native state energetics can be quantified by using the so-called Φ_M analysis²⁷, equation (1). A detailed description of this parameter, defined below, can be found elsewhere^{27,28}.

$$\Phi_M = \frac{\Delta G_{\text{mut}}^\ddagger - \Delta G_{\text{wt}}^\ddagger}{\Delta G_{\text{mut}} - \Delta G_{\text{wt}}} \quad (1)$$

A-to-E mutants are exceptionally well suited to Φ_M analyses, in that they specifically perturb one interaction (H-bonds) without creating a cavity, introducing new steric interactions or altering conformational preferences. The Φ_M value for each A-to-E mutant indicates how important the corresponding H-bond is to the TS of the PIN WW domain. A Φ_M value of 0 indicates that the H-bond is not formed in the TS, whereas a Φ_M value of 1 indicates that the H-bond is fully engaged in the TS²⁷. The interpretation of Φ_M values between 0 and 1 is less clear, but they can be taken to indicate the average extent of H-bond formation in the TS for proteins that have minimal frustration in their folding landscape²⁸.

Values of Φ_M were measured for all of the A-to-E mutants of the PIN WW domain with a midpoint of thermal denaturation temperature (T_m) over 35 °C (see Table 1). These are shown in Fig. 3, along with Φ_M values measured previously for side chain mutants at the same positions, which are included for comparison. The A-to-E mutations in loop 1 (S16σ, R17ρ and S19σ) have the highest Φ_M

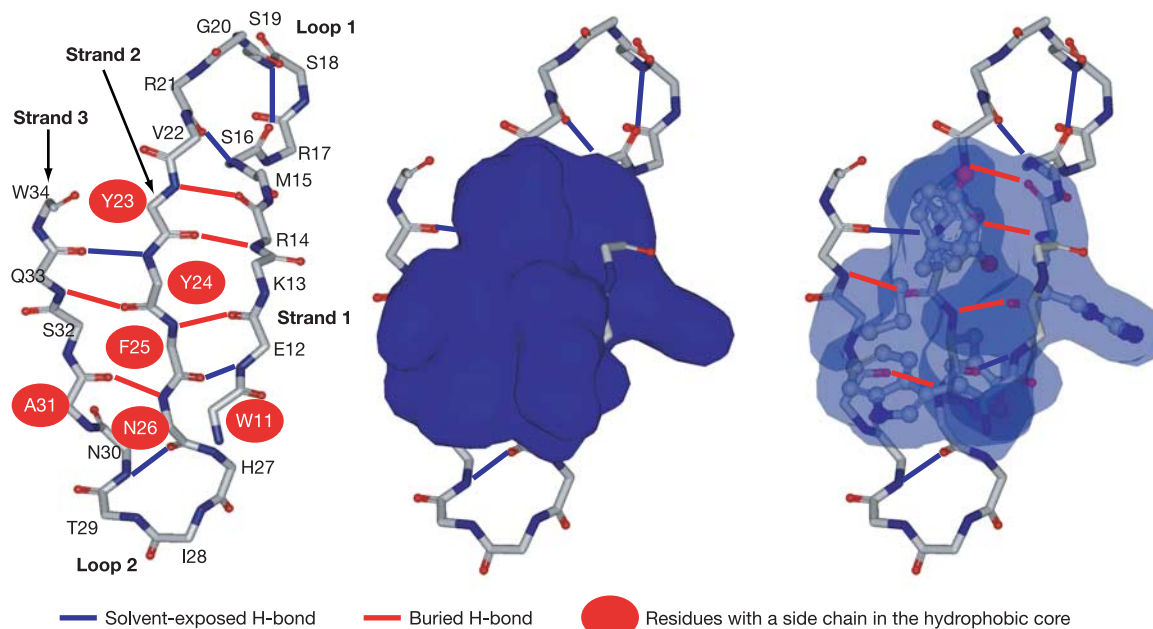


Figure 1 Structures of the PIN WW domain. The H-bond locations are shown relative to the hydrophobic core, which is indicated by the transparent and dark blue shading.

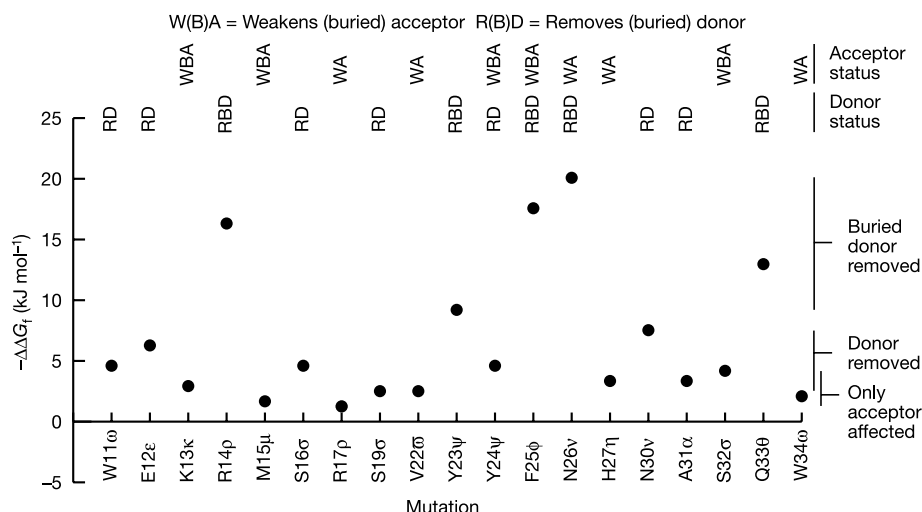


Figure 2 Plot of $-\Delta\Delta G_f$ at 2 °C for the amide-to-ester (A-to-E) mutants versus mutation location. $\Delta\Delta G_f = (\Delta G_f \text{ (wild type)} - \Delta G_f \text{ (A-to-E mutant)})$. Whether the A-to-E mutation affects the H-bond donor and/or acceptor is indicated along the top of the plot. The values

of $-\Delta\Delta G_f$ fall into ranges, indicated on the right, depending on whether the mutation weakens an acceptor, removes a donor or removes a buried donor.

values (0.70, 0.78 and 0.83, respectively), indicating that the backbone H-bonds in loop 1 are almost fully formed in the TS. The Φ_M values for residues in the strands (K13 κ , V22 τ and S32 σ) and loop 2 (H27 η) are lower than those for residues in loop 1 but are still substantial (average 0.43, range 0.28–0.51), indicating that the H-bond-mediated secondary structure in these parts of the PIN WW domain exists in the TS, but is less well developed than in loop 1. The Φ_M value for W34 ω , which is near the end of strand 3, is roughly 0, indicating that this part of the PIN WW domain is essentially unstructured in the TS. As the temperature is increased from 50 to 70 °C, the TS becomes more structured, particularly around S32 (the Φ_M of S32 σ increases from 0.51 at 50 °C to 1.0 at 70 °C; see Supplementary Information for a plot of Φ_M versus temperature for each residue).

The above observations are broadly consistent with the conclusions drawn from side chain mutant Φ_M values¹⁷ in that residues in loop 1 dominate the TS, whereas residues in the hydrophobic core dominate thermodynamic stability. There is, however, an important difference. The Φ_M values at 50 °C for the A-to-E mutants from position 16 to 22 of the PIN WW domain (S16 σ , R17 ρ , S19 σ and V22 τ) are smaller than those for side chain mutants in the same segment (S16T, S18G, S19G and V22A), which exceed 1 (Fig. 3). According to equation (1), Φ_M values greater than 1 indicate that a residue makes an interaction that stabilizes the TS more than the native state. Because the side chain Φ_M values in loop 1 are greater than 1, whereas the A-to-E Φ_M values are close to, but still less than, 1, it can be inferred, first, that the side chain hydroxyl groups of S16,

S18 and S19 are responsible for tertiary structural interactions that stabilize the TS and, second, that these side chain interactions exist in a TS with native-like secondary structure from the carbonyl of S16 to the NH group of S19.

A-to-E mutations are an incisive tool for probing the influence of backbone H-bonding on folding kinetics and thermodynamics, because they perturb H-bonding without affecting conformational preferences or side chain interactions. For the PIN WW domain, A-to-E mutations have allowed us to substantiate some results obtained from side chain mutants. In particular, we have corroborated that there are two groups of residues in this domain: those that determine the stability of the folded PIN WW domain (such as R14, Y23, F25, N26 and Q33), and those that determine the rate of folding (such as S16, R17 and S19). A-to-E mutations have also allowed us to extend our understanding of PIN WW domain folding in a way that would be impossible by side chain mutagenesis. In particular, they have allowed us to isolate and distinguish the effects of side chain interactions and backbone H-bonding on TS energies. This is particularly marked in loop 1, where the Φ_M values from A-to-E mutants show that native-like H-bond-mediated secondary structure is formed in the folding TS, and that side chain/side chain interactions occur within this framework that stabilize the TS more than the native state.

Combining the data from side chain and A-to-E mutagenesis yields the following, highly detailed picture of PIN WW domain folding. At 50 °C, the TS has native-like H-bond-mediated secondary structure in loop 1. The TS is stabilized by interactions involving

Table 1 Summary of thermodynamic and kinetic data of the A-to-E PIN WW mutants stable enough for laser T-jump measurements.

	Thermodynamics		Kinetics		
	T_m (°C)	ΔG_f , 50 °C (kJ mol ⁻¹)*	$t_{1/2, \text{folding}}$, 50 °C (μs)	$\Delta G_f^\ddagger$, 50 °C (kJ mol ⁻¹)‡	Φ_M , 50 °C (±s.e.m.)
Wild type	59.0	-2.6	110	21.3	—
K13 κ	46.4	1.6	190	23.4	0.50 ± 0.05
S16 σ	42.2	3.2	420	25.3	0.70 ± 0.05
R17 ρ	49.1	0.5	210	23.6	0.78 ± 0.11
S19 σ	38.4	5.9	1,100	28.4	0.83 ± 0.04
V22 τ	56.7	-2.0	120	21.5	0.42 ± 0.31
H27 η	38.7	4.8	150	23.4	0.28 ± 0.03
S32 σ	41.5	4.8	290	25.0	0.51 ± 0.03
W34 ω	49.5	0.0	80	21.0	-0.12 ± 0.02

* ΔG_f , the free energy of folding, is the free energy of the native state relative to the unfolded state at 50 °C (determined by thermal denaturation); standard errors of the mean in ΔG_f are of the order of ±0.3 kJ mol⁻¹.

‡ $\Delta G_f^\ddagger$, the activation barrier to folding, is the free energy of the TS relative to the unfolded state at 50 °C; s.e.m. in $\Delta G_f^\ddagger$ are on the order of ±0.05 to 0.1 kJ mol⁻¹.

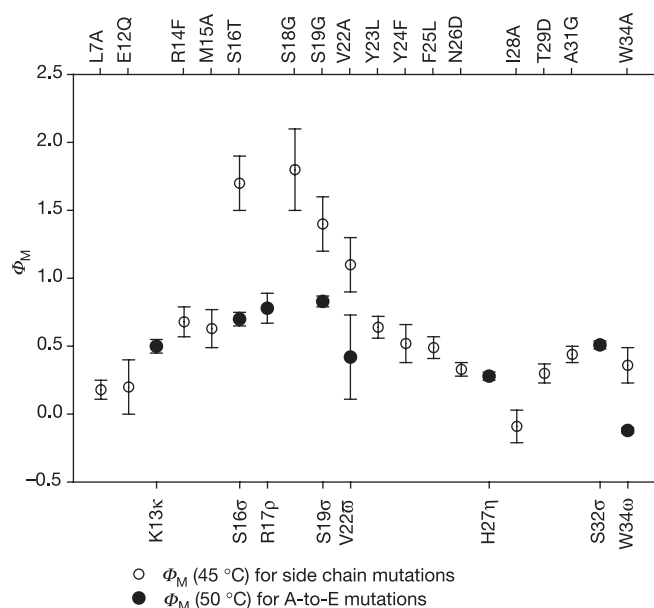


Figure 3 Plot of Φ_M versus mutation location. Filled black circles indicate the Φ_M values of A-to-E mutations at 50 °C, with the mutation given on the bottom axis. Open circles represent the Φ_M values of side chain mutations at 45 °C, with the mutation identified on the top axis. Error bars represent s.e.m.

the S16, S18 and S19 side chains, but these interactions become weaker, or non-existent, in the native state. The strands and loop 2 have some H-bond-mediated secondary structure in the TS, but are significantly less native-like than loop 1, as judged by their lower Φ_M values. The end of the third strand (around residue W34) is almost completely unstructured. At higher temperatures, the TS becomes more structured in general, and particularly around S32. Finally, the stability of the folded structure depends on the formation of a hydrophobic core, and on the formation of H-bonds by the backbone amides within this hydrophobic core. The detail attained in this study of β -sheet folding, made possible through the combined use of A-to-E and side chain mutagenesis data, should stimulate further theoretical and experimental efforts to chart the trajectory of PIN WW domain folding at an atomic level. □

Methods

Preparation of A-to-E mutants of the PIN WW domain

The wild-type and A-to-E variants of the PIN WW domain were chemically synthesized by using a solid-phase Boc strategy as described^{12,29}. Although a few of the α -hydroxy acids are commercially available, many of the protected α -hydroxy acids required were synthesized by methodology that was developed or optimized in our laboratory²⁵. We purified the A-to-E PIN WW domain variants by reverse-phase HPLC and then subjected them to size-exclusion chromatography as described elsewhere (S.D., P.E.D. and J.W.K., unpublished results). Depsipeptide PIN WW domain variants with a moderately labile ester bond were stored as lyophilized powders that were dissolved in deionized water and diluted with 20 mM sodium phosphate buffer (pH 7.0) to the desired concentration, determined by ultraviolet absorption (extinction coefficient, $\epsilon_{280} = 13,940 \text{ M}^{-1} \text{ cm}^{-1}$).

Assessing native state stability

A comprehensive description of the equilibrium properties of A-to-E variants of the PIN WW domain will be presented elsewhere (S.D., P.E.D. and J.W.K., unpublished results). In brief, steady-state thermal denaturation curves were recorded by far-ultraviolet circular dichroism spectroscopy (227 nm) in 20 mM sodium phosphate buffer (pH 7.0) on a 202SF circular dichroism spectrometer (AVIV), equipped with a peltier temperature controller. The temperature was raised in increments of 2 °C from 2 to 98 °C. The protein solution was allowed to equilibrate for 90 s at each temperature before the signal at 227 nm was averaged for 30 s. The thermal denaturation traces were fit to a two-state folding model with the free energy of folding expressed as a Taylor expansion about T_m , $\Delta G_f(T) = \Delta G_0 + \Delta G_1(T - T_m) + \Delta G_2(T - T_m)^2$, as described^{16,17}. The values for ΔG_0 , ΔG_1 and ΔG_2 for each mutant are tabulated in the Supplementary Information. Values for $K_{eq}(T)$ were then calculated from $\Delta G_f(T)$. Guanidine hydrochloride denaturation curves were recorded and used to extract ΔG_f for Y23 ψ , F25 ϕ and N26 ν as described elsewhere (S.D., P.E.D. and J.W.K., unpublished results).

T-jump kinetic studies

The fast folding and unfolding rates of PIN WW domains were measured by a laser T-jump apparatus³⁰ that was previously applied to PIN WW domain studies^{16,17,20}. The equilibrium is perturbed on a sudden increase in temperature (roughly 12 °C), achieved and controlled by two counter-propagating heating pulses from a Q-switched Nd:YAG laser, which was Raman-shifted from 1.064 to 1.54 μm in 2 MPa of methane. The rate constants (RCs) of relaxation back to equilibrium, the sum of folding (k_f) and unfolding (k_u) RCs, after a nanosecond laser T-jump were determined from tryptophan fluorescence lifetime decays (excitation every 14 ns by an ultraviolet pulse train centred at 290 nm; data were digitized with a resolution of 500 ps). Fused silica sample cuvettes with path lengths of 0.3 mm were used for all measurements and tryptophan calibrations. We carried out tryptophan calibrations at the beginning of a measurement day and sometimes later in the day to verify T-jump consistency. The fluorescence decay trace was then fit to a two-state model to obtain the relaxation RC, $k_{obs}(T)$, which is the sum of the folding, $k_f(T)$, and unfolding $k_u(T)$, RCs. The measured equilibrium constant, $K_{eq}(T) = k_f(T)/k_u(T)$, from steady-state (thermodynamic) studies and the relaxation RC, $k_{obs}(T) = k_f(T) + k_u(T)$, from kinetics studies enable the folding RC, $k_f(T)$, and unfolding RC, $k_u(T)$, to be calculated. We obtained the free energies of activation for folding from the folding RCs using a Kramers model as described¹⁷ (see Supplementary Information). The parameters that define the temperature dependence of $\Delta G_f^\ddagger$ are also tabulated in the Supplementary Information. The laser T-jump method requires that the PIN WW domain is partially folded over the temperature range of 30–80 °C (ref. 17). At lower temperatures measurements are complicated by noise, whereas at temperatures above 80 °C the ester linkage becomes hydrolytically unstable.

Received 12 January; accepted 29 April 2004; doi:10.1038/nature02611.

- Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).
- Honig, B. & Yang, A. S. Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27–58 (1995).
- Myers, J. K. & Pace, C. N. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **71**, 2033–2039 (1996).
- Sudol, M. Structure and function of the WW domain. *Prog. Biophys. Mol. Biol.* **65**, 113–132 (1996).
- Thorson, J. S., Chapman, E. & Schultz, P. G. Analysis of hydrogen bonding strengths in proteins using unnatural amino acids. *J. Am. Chem. Soc.* **117**, 9361–9362 (1995).
- Chapman, E., Thorson, J. S. & Schultz, P. G. Mutational analysis of backbone hydrogen bonds in staphylococcal nuclease. *J. Am. Chem. Soc.* **119**, 7151–7152 (1997).
- Koh, J. T., Cornish, V. W. & Schultz, P. G. An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **36**, 11314–11322 (1997).
- Shin, L., Ting, A. Y. & Schultz, P. G. Analysis of backbone hydrogen bonding in a β -turn of staphylococcal nuclease. *J. Am. Chem. Soc.* **119**, 12667–12668 (1997).
- Haq, T. S., Little, J. C. & Gellman, S. H. Stereochemical requirements for β -hairpin formation: model studies with four-residue peptides and decapeptides. *J. Am. Chem. Soc.* **118**, 6975–6985 (1996).
- Wales, T. E. & Fitzgerald, M. C. The energetic contribution of backbone-backbone hydrogen bonds to the thermodynamic stability of a hyperstable p22 Arc repressor mutant. *J. Am. Chem. Soc.* **123**, 7709–7710 (2001).
- Silinski, P. & Fitzgerald, M. C. Comparative analysis of two different amide-to-ester bond mutations in the β -sheet of 4-oxalocrotonate tautomerase. *Biochemistry* **42**, 6620–6630 (2003).
- Beligere, G. S. & Dawson, P. E. Design, synthesis, and characterization of 4-ester C12, a model for backbone hydrogen bonding in protein α -helices. *J. Am. Chem. Soc.* **122**, 12079–12082 (2000).
- Blankenship, J. W., Balambika, R. & Dawson, P. E. Probing backbone hydrogen bonds in the hydrophobic core of GCN4. *Biochemistry* **41**, 15676–15684 (2002).
- Chen, H. I. et al. Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J. Biol. Chem.* **272**, 17070–17077 (1997).
- Koepf, E. K., Pettrassi, H. M., Sudol, M. & Kelly, J. W. WW: an isolated three-stranded antiparallel β -sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* **8**, 841–853 (1999).
- Crane, J. C., Koepf, E. K., Kelly, J. W. & Gruebele, M. Mapping the transition state of the WW domain β -sheet. *J. Mol. Biol.* **298**, 283–292 (2000).
- Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. The folding mechanism of a β -sheet: the WW domain. *J. Mol. Biol.* **311**, 373–393 (2001).
- Kowalski, J. A., Liu, K. & Kelly, J. W. NMR solution structure of the isolated apo Pin1 WW domain: comparison to the X-ray crystal structures of Pin1. *Biopolymers* **63**, 111–121 (2002).
- Kaul, R., Deechongkit, S. & Kelly, J. W. Synthesis of a negatively charged dibenzofuran-based β -turn mimetic and its incorporation into the WW miniprotein-enhanced solubility without a loss of thermodynamic stability. *J. Am. Chem. Soc.* **124**, 11900–11907 (2002).
- Nguyen, H., Jager, M., Moretto, A., Gruebele, M. & Kelly, J. W. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl Acad. Sci. USA* **100**, 3948–3953 (2003).
- Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H. & Fersht, A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl Acad. Sci. USA* **98**, 13002–13007 (2001).
- Ferguson, N. et al. Using flexible loop mimetics to extend Φ -value analysis to secondary structure interactions. *Proc. Natl Acad. Sci. USA* **98**, 13008–13013 (2001).
- Joseph, J. D., Yeh, E. S., Swenson, K. I., Means, A. R. & Winkler, K. E. The peptidyl-prolyl isomerase Pin1. *Prog. Cell Cycle Res.* **5**, 477–487 (2003).
- Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**, 875–886 (1997).
- Deechongkit, S., You, S.-L. & Kelly, J. W. Synthesis of side chain protected chiral α -hydroxy acids for Boc solid phase peptide synthesis. *Org. Lett.* **6**, 497–500 (2004).
- Baskakov, I. & Bolen, D. W. Forcing thermodynamically unfolded proteins to fold. *J. Biol. Chem.* **273**, 4831–4834 (1998).
- Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782 (1992).
- Nymeyer, H., Socci, N. D. & Onuchic, J. N. Landscape approaches for determining the ensemble of

folding transition states: success and failure hinge on the degree of frustration. *Proc. Natl Acad. Sci. USA* **97**, 634–639 (2000).

29. Bramson, H. N., Thomas, N. E. & Kaiser, E. T. The use of *N*-methylated peptides and depsipeptides to probe the binding of heptapeptide substrates to cAMP-dependent protein kinase. *J. Biol. Chem.* **260**, 5452–5457 (1985).
30. Grubele, M., Sabelko, J., Ballew, R. & Ervin, J. Laser temperature jump induced protein refolding. *Acc. Chem. Res.* **31**, 699–707 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. You, J. Blankenship and R. Balambika for discussions. We acknowledge financial support from the NIH, The Skaggs Institute for Chemical Biology, the Lita Annenberg Hazen Foundation and the Norton B. Gilula Fellowship (to S.D.). M.G. and H.N. were supported by a grant from the NSF. P.E.D. was supported by the NIH and the Alfred P. Sloan Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.W.K. (jkelly@scripps.edu).

corrigendum

Global-scale temperature patterns and climate forcing over the past six centuries

Michael E. Mann, Raymond S. Bradley & Malcolm K. Hughes

Nature **392**, 779–787 (1998).

It has been drawn to our attention (S. McIntyre and R. McKittrick) that the listing of the ‘proxy’ data set in the Supplementary Information published with this Article contained several errors. In Table 1 we provide a list of the records that were either mistakenly included in the Supplementary Information, or mistakenly left out. A small number of other corrections of the original listing include (see Table 1) corrections of the citations originally provided, or corrections of the start years for certain series.

The full, corrected listing of the data is supplied as Supplementary Information to this corrigendum. Also provided as Supplementary Information are a documented archive of the complete data (instrumental and ‘proxy’ climate series) used in our original

Table 1 Errors in ‘proxy’ data set listing in ref. 1

Series (34) listed in original Supplementary Information but not used in ref. 1.*

FRAN003; ITAL015 and ITAL015X; SPAI026 and SPAI047; NEWZ056; ARGE030, ARGE060 and ARGE065; CHIL015, CHIL016, CHIL017 and CHIL018; AK006 and AK006X; CA070; CANA053, CANA053X, CANA096, CANA096X, CANA099, CANA106 and CANA110; WA019, WA025, WA027, WA033, WA039, WA041, WA071, WA074, WA086, WA088 and WA091; VAGANOV55

Series (2) used in ref. 1 but not listed in original Supplementary Information

Unpublished Southwest US/Mexico Density series (D. W. Stahle, personal communication)

Unpublished Southwest US/Mexico Latewood Width series (D. W. Stahle, personal communication)

Additional minor corrections

(1) The Central England and Central European temperature records used by ref. 1 were the summer season versions of these series as used by ref. 2.

(2) The ‘long instrumental’ series used in ref. 1 are station temperature and precipitation station data from the NOAA Climate Data centre gridded at 5° latitude/longitude resolution.

(3) The start year for the ‘Central Europe’ series of ref. 1 is AD 1525.

(4) The ‘Western North America Dendro density’ series used in ref. 1 should properly be attributed to ref. 3.

(5) The Stahle *et al.* Southwestern/Mexico late wood width and maximum density data used in ref. 1 should properly be attributed to ref. 4 (the formal reference was not available at the time of ref. 1), or, in two cases, unpublished data (D. W. Stahle, personal communication).

(6) For one of the 12 ‘Northern Treeline’ records of Jacoby *et al.* used in ref. 1 (the ‘St Anne River’ series), the values used for AD 1400–03 were equal to the value for the first available year (AD 1404).

* These series, all of which come from the International Tree Ring Data Bank (ITRDB), met all the tests used for screening of the ITRDB data used in ref. 1 (see ref. 5), except one—namely, that in 1997, either it could not be ascertained by the authors how these series had been standardized by the original contributors, or it was known that the series had been aggressively standardized, removing multidecadal to century-scale fluctuations.

study, and an expanded description of the methodological details of our original study.

None of these errors affect our previously published results¹. □

1. Mann, M. E., Bradley, R. S. & Hughes, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).
2. Bradley, R. S. & Jones, P. D. “Little ice age” summer temperature variations: their nature and relevance to recent global warming trends. *The Holocene* **3**, 367–376 (1993).
3. Briffa, K. R. *et al.* Fennoscandian summers from AD500: temperature changes on short and long timescales. *Clim. Dyn.* **7**, 111–119 (1992).
4. Stahle, D. W. *et al.* Experimental dendroclimatic reconstruction of the Southern Oscillation. *Bull. Am. Meteorol. Soc.* **79**, 2137–2152 (1998).
5. Mann, M. E. *et al.* Global temperature patterns in past centuries: An interactive presentation. *Earth Inter.* **4-4**, 1–29 (2000).

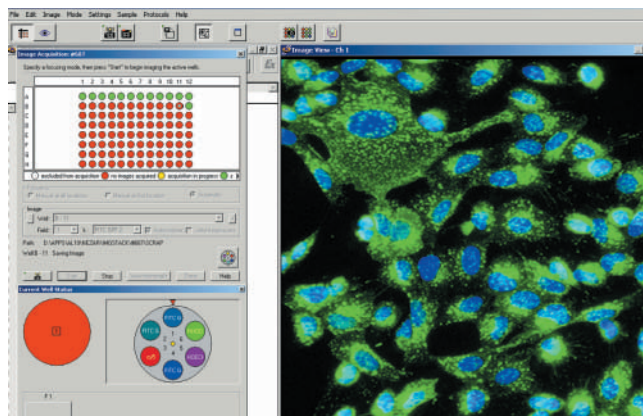
Supplementary Information accompanies this corrigendum on www.nature.com/nature.

The leading edge

Synthesizing and selecting the lead compounds that will become the drugs of the future are the heart of drug discovery. Tim Chapman goes back to basics.

Over the past decade, high-throughput screening (HTS) has revolutionized the process of drug discovery. Automation and miniaturization have allowed biotech companies to screen hundreds of thousands of compounds against disease targets, in the hope of identifying the precious few that can be developed into commercial, patentable medicines. To use a fashionable analogy, it's like speed dating for medicines.

But this approach of simply testing more and more compounds at faster and faster rates is not as productive as the industry had hoped. Because the earliest stages of the discovery process are relatively unselective, the vast majority of early hits fail to progress, either because of the attrition caused by medical or chemical failings discovered only in the later stages of development, or because the flow of potential leads is still too overwhelming. Drug-discovery researchers, and the companies that supply their technological needs, are now concentrating on improving both the quality of the compounds entering the first screening and the



Looking in: output from the IN Cell Analyzer 1000.

value of the assays being done. Rather than bigger and faster, the focus is on doing things earlier and smarter.

The demand for high-throughput screening began in the early 1990s, when drug companies began testing the landslide of compounds created by combinatorial chemistry against an increasing number of disease targets identified by genomic research. "What we then realized was that the bottleneck, rather than being in the set-up and running of the assay, moved to reading the assays themselves," says John Anson,

vice-president of development at GE Healthcare Bio-Sciences (the former Amersham Biosciences), based in Little Chalfont, UK.

Automated assays demanded the use of plates with arrays of wells, and early systems measured each well individually. As plate densities increased up to 1,536 wells per plate, complex camera systems became available that could read the whole plate simultaneously. GE Healthcare's high-throughput screening platform, LEADseeker, is now in its fourth iteration and offers a range of imaging modes within the single instrument, allowing richer measurements.

"What's interesting about high-throughput screening is that there's no dominant way of measuring an assay," Anson notes. "Customers can use luminescence, radioactivity, or lots of different fluorescence modalities. There's not one dominant modality, so we had to focus on sensitivity and assay performance as well as speed." In fluorescence mode, LEADseeker reads a plate in about a minute, allowing throughputs of up to half-a-million assays per day. "It was that ability to

GE HEALTHCARE

MINING THE PROTEOME

Most automated approaches to drug discovery are based on testing thousands of compounds against a single disease target. The platform developed by Serenex, a discovery company in Durham, North Carolina, takes the opposite approach by testing drug candidates against thousands of protein targets simultaneously. This technique, which the firm calls proteome mining, promises to deliver unique information about interactions between compounds and targets.

"In contrast to profiling compounds against a single target, what we're doing is profiling one compound against all targets in an entire sub-proteome," says Steven Hall, the company's senior vice-president of R&D. "We get extremely broad selectivity and information."

Serenex is focusing on the purine-binding proteins within the human proteome. This group of some 2,000 proteins includes protein kinases and many other clinically important proteins that use purine nucleotides such as adenosine triphosphate as co-substrates or co-factors. A potential

drug candidate can be tested against all 2,000 in a single step.

The approach depends on a patented affinity medium of small beads to which many molecules of a purine nucleotide such as ATP are tethered. When a cell or tissue extract is mixed with this medium, the tethered molecules ensnare its purine-binding proteins. Unbound proteins are then washed away, and a drug compound introduced. If the compound has an affinity for any of the bound proteins, it will displace them from the tethered purines. A selective drug will displace just one protein, whereas less-selective compounds will displace many different proteins. The displaced proteins can then be collected and identified by mass spectrometry and bioinformatics.

The aim is to help medicinal chemists make a more informed decision about which compounds to move forward to lead optimization. Larger-scale screening of large compound libraries against sub-proteome sets of proteins will provide a very useful and information-rich matrix of compound-target interactions, Hall notes.

"We feel that we will get to the point where we can develop some fundamental understandings of what kind of substructures will lead to effects on what type of target," he says. "It's only by screening large libraries of targets against large numbers of compounds that the informatics community will be able to get the connection between structure and protein affinity."

T.C.



Taking a molecular view.

TECAN

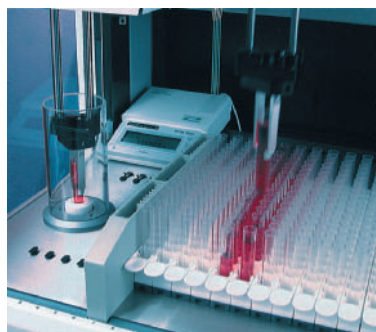
run things at this throughput that helped us realize where the next bottleneck would be, which is turning these hits into lead compounds," says Anson.

Into the cell

The vast majority of compounds identified as hits in screening fail in the later stages of lead development, usually because they are biologically unsuitable for development as a medicine. It makes sense to find out more about the biological workings of a hit compound as early as possible, before you spend time and money on further development.

"One area where we are seeing continuing demand for products in the early stages of drug discovery involves cell-based assays," says Marc Feiglin, biopharma technology officer at life-sciences equipment supplier Tecan, based in Männedorf, Switzerland. Tecan's Cellerity system provides an automated system for high-throughput cell-culture production and assay preparation.

"One of the reasons people are using cell-based assays is they're starting to get earlier indications of problems that may occur later on," says Anson. GE Healthcare's IN Cell Analyzer platforms offer real-time high-resolution cellular imaging, allowing researchers to construct high-throughput intracellular assays that were previously impossible (see 'Proteins on the move', below).



Automated analytics.

based in Pittsburgh, Pennsylvania. "You can screen millions of compounds very quickly, but the challenge for pharma is getting enough information out of that screening to make considered decisions to move forwards," he says.

Cellomics' high-content screening platforms such as the ArrayScan and KineticScan readers use fluorescence probes to obtain richer measurements of cellular activity. "The point is that you're measuring more than one thing that's going on, and analysing images in an automated fashion to get quantitative data," Collins says. "You can address biology you could never have understood before, things like morphology screening that you just can't do if you've smashed the cell up and put it in a test tube."

Getting the knowledge

With both throughputs and the richness of assay measurements increasing, making sense of the vast amounts of data generated can be a bottleneck. Informatics or knowl-

edge-management systems are a necessity, and many platform providers offer integrated software packages.

"When you are generating multiple images and need to get a quantitative image out quickly, you're not going to do that by running millions of spreadsheets or looking at millions of images," says Collins. "What our informatics package does is automate that whole process of linking data together and linking images with numbers and storing that in a very robust database. The image is always there as a final answer, but once the instrument has done its job you're looking at numbers, which is a sight easier than looking at 1,000 images."

Informatics can also help reduce attrition by improving the selection of hits to take into further development. "By integrating our synthetic chemistry work with a knowledge-management approach, we've demonstrated that we can improve the interaction profile of a customer's project by increasing the number of lead series they have for each target, which gives them more shots at the goal," says Ed Hodgkin, vice-president for business development at discovery services group Tripos, based in St Louis, Missouri.

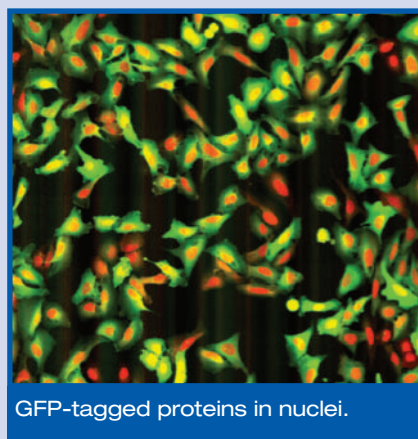
In a collaboration with New York-based pharmaceutical giant Pfizer, Tripos has developed the SARNavigator suite of tools for analysing high-throughput screening results. The package allows researchers to examine structure-activity relations (SAR) quantitatively to help determine which hits should go forward into the lead identification and optimization phase. "That combination of SAR-Navigator and the chemical compounds we

PROTEINS ON THE MOVE

BIOMAGE

Intracellular signalling pathways are involved in virtually every disease process, and cell-surface receptor proteins that initiate these pathways are favoured drug targets (see *Nature* **428**, 225–233; 2004). In the intracellular pathways leading from these receptors, signalling proteins move within the cell — from the cytoplasm into the nucleus, for example. Testing the effects of potential drug compounds against these intracellular protein movements has, unsurprisingly, proved much harder than testing against proteins at the surface of the cell. The ability to do so promises to find drugs with new modes of action and greater selectivity.

A leader in this field is Biomage in Copenhagen, Denmark, which focuses on green fluorescent proteins (GFP) as a tool in drug discovery. Biomage's proprietary Redistribution technology uses genetically encoded GFP tags to track protein translocation in living cells, and so identify compounds that inhibit onward transmission of the signal. "It's pretty clear now that every signalling pathway has components that must move in the cell for the signal interaction to occur," says Len Pagliaro, vice-president for business development at Biomage. "If you take these targets and



GFP-tagged proteins in nuclei.

track them, you can measure movement in response to potential stimulators and agonists, and have the potential to find compounds that inhibit that movement."

The firm deploys GE Healthcare's IN Cell Analyzer to carry out high-throughput imaging for screening and analysis of the effects of members of its own library of over 250,000 compounds. Such intracellular assays provide the opportunity to discover drugs that interact with target proteins in new ways — by targeting protein localization domains rather than catalytic domains, for instance. Studying living cells also creates more opportunities for serendipitous discoveries.

"Obviously, cell-based assays are inherently a somewhat dirty system biologically, and that has both benefits and liabilities," Pagliaro notes. "The

benefit is that in principle it's more like the pathway you're going to see *in vivo*. The liability is that when you get a hit, you don't know that the compound is acting on the target that you want to see it acting on. We like that aspect, because that can deliver very interesting progressible hits that we would not have found otherwise."

T.C.

provide is a very powerful combination in going from a hit to a lead," Hodgkin says.

Applying informatics along the discovery process can reduce the time taken in each phase, both by exploiting data mining to reduce the attrition of candidate compounds and by managing the logistics of the whole process. Benchmarking exercises by Tripos and Pfizer have found that the time taken to go from initial screening to a preclinical drug candidate can be reduced to 9–18 months from an industry standard of 2½–3 years.

Combining data-mining techniques with high-precision screening and library selection can offer dramatic improvements in productivity. Drug-discovery company Amphora, at Research Triangle Park, North Carolina, focuses on finding small-molecule therapeutics for big-pharma partners such as Strasbourg-based Aventis. "The quality of the data produced at Amphora has a direct effect on drug-discovery efficiency," says Jeff Riley, vice-president of corporate development. Structure–activity relation modelling allows the firm's chemists to move directly from primary screening to medicinal chemistry stage. "This enables Amphora to efficiently and repeatedly go from target to lead in six months," Riley notes. "This is true predictable drug discovery."

Library science

Amphora also prides itself on the selectivity and purity of its compound library. "The old computer software adage, garbage in garbage out definitely applies to HTS," says Riley. "The higher the purity of the library and the more that is known about the quan-

titation, the higher the quality of the data."

Improving the quality of the compounds going into initial screening is perhaps the key factor in improving the usefulness of the hits coming out. Researchers are increasingly realizing the waste involved in simply screening vast numbers of compounds without sufficient preselection.

"I do not believe that you can solve the problem of finding drugs through a numbers game," says Stephen Hill, chief executive of biotech firm ArQule, based in Woburn, Massachusetts. "The problem with high-throughput screening is there's almost no end to it, because the number of compounds you could potentially use is 10^{60} . If you have an ocean to choose from, it doesn't make much difference whether you sample a spoonful or a cupful—it's still not representative."

ArQule has developed a proprietary method for producing compounds with the best balance of drug-like properties, or optimal chemical entity (OCE) compounds, which it uses to offer high-quality library design and production to research partners such as Pfizer.

"The goal is to increase the likelihood that Pfizer will be able to generate from primary screening more interesting hits against a broad range of its targets," Hill says. "Because of the way we've designed those compounds, it's more likely the company will be able to optimize within that chemical space a compound with all the drug-like characteristics to make a medicine."

The OCE approach uses predictive models and absorption, distribution, metabolism and excretion (ADME) assays as early as pos-

sible, and looks for the best balance of drug-like characteristics at every step rather than simply prioritizing compounds with the highest affinity for the target. "What we try to do is think about all these parameters in parallel at every stage of the process, starting from library design and prediction through each iterative step," Hill says. "When you're trying to do a multiparameter optimization, trying to trade off up to a dozen chemistries, that's when you need the power of computational tools backed up by relatively quick and affordable experimental capability."

Producing high-quality, targeted compound libraries has become an industry in itself. One of the largest players is Chemical Diversity Labs (CDL), based in San Diego, California, which produces some 200,000 compounds in 500 new library systems per year. "There is a greater awareness of the problems associated with the failure of leads late in drug development, and the emphasis has shifted to their quality and survivability in the development process," says CDL president Nikolay Savchuk. "Customers want screening libraries with new chemistry, preferably beyond the scope of current patents and literature references, but also with a higher probability of hitting the target."

CDL's targeted libraries are derived from annotated databases that unify knowledge about targets, ligands and their relationships. Compounds are also ranked by factors such as novelty and ease of synthesis.

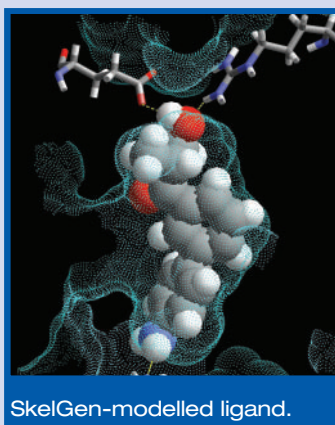
Entirely new compounds designed by molecular modelling (see 'Starting from scratch', below) create extra challenges for chemical synthesis, but the results can justify

STARTING FROM SCRATCH

DE NOVO PHARMACEUTICALS

Computational methods are being applied across the drug-development process from the earliest stages of drug design. "We're seeing a lot of companies who don't want to spend time and money on screening existing compound libraries," says Bart Wuurman, chief executive officer at De Novo Pharmaceuticals, a leading *in silico* drug-design company in Cambridge, UK. "The traditional approach is looking for a needle in a haystack—what we do is create a needle."

Founded in 1999 as a spin-out from the University of Cambridge's drug-design group, De Novo last year refocused itself from in-house drug discovery to meeting the growing demand for its proprietary computational platforms. De Novo's core platform is a suite of structure-generating programs called SkelGen. The core algorithm generates new molecular structures by the random assembly of small molecular fragments. This stochastic approach rapidly searches the whole of the chemical space of potential drug compounds, explains chief scientific officer Philip Dean. The search is constrained by any information available about the structure of the protein target from X-ray crystallography. In cases where little is known about the target, information about the structure of a known ligand that affects the



SkelGen-modelled ligand.

target's activity can be exploited. The firm's Quasi2 program uses data on the known active ligands to create a negative image of the target site, which can then be input into the SkelGen algorithm to find new patentable compounds with a similar action.

"The unique technology is the ability to create novel chemotypes for any set of constraints," Dean says. "We think there's no one else out there who can do *de novo* design from just an active compound." The firm's researchers have also developed a method for accommodating the flexibility of the protein target, which structure-based molecular design has previously failed to do.

De Novo's platform takes into account the need to convert *in silico* designs into actual chemicals. "We have an algorithm that post-processes the SkelGen

structures to prioritize those structures that are readily synthesizable in the smallest number of steps," Dean says. "Not only are the compounds scored for the fit with the site but also for ease of synthesis." In collaborations with Roche, over 90% of output compounds have been relatively simple for chemists to synthesize. "Basically it's our intent to get the designed molecules into wet chemistry as quickly as possible," Wuurman says. **T.C.**

the effort. "There should be a good rationale for making molecules, and *de novo* design provides such a rationale," says Savchuk. "To be successful, *de novo* design needs to be innovative but cost-efficient. As with other types of synthesis, *de novo* design has a better chance of success when screening libraries instead of single compounds are produced."

Specifically targeted compound libraries will also allow the use of more complex assays earlier in the discovery process. "If you're using a much more focused library, you can address much more complex biology at the first screen," says Anson. "I'm quite excited about moving in this direction because lots of our cell-based assays will be very amenable to high-information content screens against these focused libraries."

Back to nature

Researchers are also looking to broaden the sources of libraries for screening. Compounds created by combinatorial chemistry have dominated discovery programmes for the past decade, but attention is now returning to the wealth of compounds derived from natural products. Part of the reason is that combinatorial techniques may, paradoxically, be limiting the diversity of compound libraries.

"People traditionally built combinatorial libraries out of small molecules that they could easily synthesize, and there's a huge overlap with the libraries that other people have," says Dirk Ullmann, head of biology at discovery services group Evotec OAI in Hamburg, Germany. "Library purity and stability is also a concern."

Natural products offer unsurpassed chemical diversity and novelty, often with better affinity characteristics and evolutionary selection for biological activity. "The big advantage is you don't meet this kind of high number attrition you meet with small-molecule chemistry," Ullmann notes. "You end up much earlier with a good lead that might even be the natural product itself."

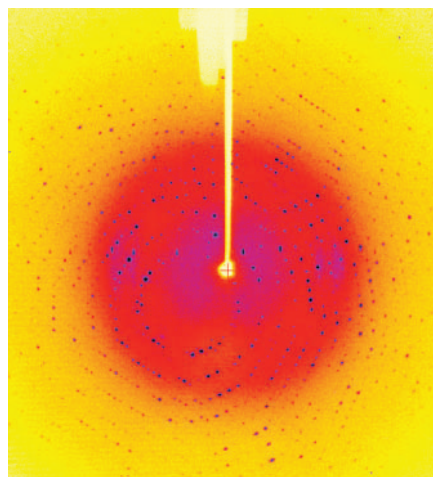
Testing natural products in a high-throughput environment is inevitably more challenging than dealing with synthetic libraries. Compounds from natural products have first to be isolated and fractionated, and creating suites of chemical analogues is more difficult.

Evotec OAI recently partnered with Biofrontera Pharmaceuticals in Leverkusen, Germany, to offer a high-throughput screening platform for one important group of natural products, those produced by bacteria and fungi, the source of many current antibiotics. The EVObioscreen platform accesses Biofrontera's largely unscreened library of fractions from some 50,000 strains of actinomycetes and fungi at ultrahigh throughputs. "It's much more unpredictable to work with plants or other organisms, so you really have

to work with microorganisms," says Ullmann.

An even more targeted supply of drug candidates may be provided by a new approach to drug discovery being piloted by a handful of companies in Europe and the United States. Fragment-based discovery uses very small molecules, termed fragments or scaffolds, as a basis for building designer compounds that may be able to tackle previously intractable disease targets.

"The whole concept of fragment-based drug discovery has gained quite a high profile in the past two or three years in the industry," says Harren Jhoti, chief scientific officer at Astex Technology in Cambridge, UK. "It's



Crystal clear: X-ray diffraction data from a protein-ligand complex.

fair to say that the hope that we all had for the high-throughput approach has perhaps not been met. One limitation has been the complexity of the compounds themselves."

Using very small molecular fragments as a foundation for building precisely targeted drug molecules allows you to keep the complexity to a minimum. The fragments in Astex's library, most of which are selected *in silico* on the basis of predicted binding efficiency, have relative molecular masses (M_r) of 100–250 and are used as the basis of compounds of M_r up to 500.

On the scaffold

Plexxikon, a scaffold-based drug-discovery company in Berkeley, California, has a library of some 20,000 scaffolds of M_r 150–300. Again, these are used as the basis of drug compounds with the smallest possible molecular masses. "What we generally find is that these leads tend to be more soluble, have better metabolic stability, and better *in vivo* ability from the very beginning," says Mike Milburn, the company's senior vice-president of research. Plexxikon applies structural screening as early as possible in the discovery process, allowing the firm's chemists to identify the most optimizable compounds as quickly as possible.

Because fragments have an inherently low binding affinity, traditional bioassays are of limited use, although Plexxikon uses bioassays in its initial screening procedure. The key to fragment-based discovery lies with X-ray crystallography.

"The real unique aspect of using X-ray crystallography is that you can detect the binding of the fragment and very precisely define the binding orientation," says Jhoti. "In the past five or seven years, there's been a huge set of developments which has allowed X-ray crystal structures to be obtained in a much more efficient manner than before." (See *Nature* **421**, 661–666; 2003.) Astex's proprietary 'Pyramid' discovery platform includes high-throughput crystallography, data collection and analysis, and can generate up to 400 protein-ligand crystal structures from X-ray data in a month, comparable to the number achieved in a year by groups using more conventional crystallography tools.

Information about the co-crystal structure of the target protein and scaffolds allows chemists to select the ones that can be most efficiently optimized for synthesis. "Having that sort of information before you start the chemistry really allows you to explore many different chemotypes for a particular target," Milburn says. "That's in contrast to high-throughput screening where you may get two or three different chemotypes based on potency that may not be the most tractable in terms of synthesis chemistry."

The bottom line

With such a diverse selection of innovative technologies and approaches being deployed at the earliest stages of the drug-discovery process, some will ultimately prove to be more productive than others. But most workers in the field agree that the current situation is not sustainable, and productivity must be improved.

"Productivity in industry is getting worse year on year," says Hill. "We know that today's drug-discovery productivity is really frighteningly bad, so continuing to do things in the way they've been done in the past is unlikely to get us where we need to be."

The good news for drug companies that have already invested heavily in high-throughput screening systems is that most of the new approaches can be added to existing systems. But to make the most of them, they will have to sweat their assets, says Anson. "It's about balancing the work, to run high-throughput factory-like operations with the recognition that we're still talking about a scientific discipline and there still has to be a certain degree of thought and planning to get a successful outcome," he says. "It's the companies that combine the ability to sweat their assets with really thoughtful design and processes that are really going to come out on top."

Tim Chapman is a freelance journalist based in Halifax, UK.

Company	Products/activity	Location	URL
Lead identification and screening			
Amphora Discovery	Lead discovery and screening	North Carolina	www.amphoracorp.com
ArQule	Chemical synthesis, lead identification	Woburn, Massachusetts	www.arqule.com
Astex Technology	Fragment-based drug discovery, computational screening, high-throughput X-ray crystallography, software for protein–ligand structure determination	Cambridge, UK	www.astex-technology.com
Chemical Abstracts Service	SciFinder chemical information database	Columbus, Ohio	www.cas.org
Chembridge	Small-molecule libraries for screening	San Diego, California	www.chembridge.com
Chemical Diversity Labs	Target collection for lead discovery, target-specific libraries, building-block libraries	San Diego, California	www.chemdiv.com
Chemspeed	Automated chemical synthesis workstations, peptide synthesizers	Augst, Switzerland	www.chemspeed.com
CombinatoRx	Lead compound identification and screening; drug discovery	Boston, Massachusetts	www.combinatorx.com
CombiPure	Chemical library services	Kegworth, UK	www.combipure.com
De Novo Pharmaceuticals	<i>In silico</i> drug-design technology	Cambridge, UK	www.denovopharma.com
Evotec OAI	Services for drug discovery: compound libraries, target assays, medicinal chemistry	Hamburg, Germany	www.evotecoi.com
Graffinity	Drug-discovery using chemical microarrays, imaging of protein–ligand interactions	Heidelberg, Germany	www.graffinity.com
Iconix	DrugMatrix databases and software for chemogenomics-based research	Mountain View, California	www.iconixpharm.com
MDL	CrossFire Beilstein, CrossFire Gmelin and other chemistry-related databases	San Leandro, California	www.beilstein.com
MDS Pharma	Services company for all aspects of drug discovery	Lincoln, Nebraska	www.mdsps.com
Plexxikon	Proprietary scaffold-based Drug Discovery platform for small-molecule inhibitors	Berkeley, California	www.plexxikon.com
Schrödinger	Molecular modelling for <i>in silico</i> screening	Portland, Oregon	www.schrodinger.com
Serenex	Screening against proteins; functional proteome fractionation; proteome mining	Durham, North Carolina	www.serenex.com
Signal Pharmaceuticals	Target and lead discovery in signalling pathways	San Diego, California	www.signalpharm.com
Tripos	Chemical libraries; modelling, pharmacophore perception and screening software	St Louis, Missouri	www.tripos.com

Cell-based assay systems and high-content screening

Aclara BioSciences	LabCard 'lab-on-a-chip' systems for biochemical and cellular analysis	Mountain View, California	www.aclara.com
Asys Hitech	Microplate readers, washers and dispensers	Eugendorf, Austria	www.asyshitech.com
Axon Instruments	ImageExpress high-throughput system for live-cell assays	Foster City, California	www.axon.com
Beckman Coulter	SAGIAN core system for high-content screening	Fullerton, California	www.beckmancoulter.com
Biomol	Reagents for signal transduction research	Plymouth Meeting, Pennsylvania	www.biomol.com
Bio-Tek	Automated microplate handling and reading instrumentation, spectrophotometers	Winooski, Vermont	www.biotek.com
BMG Labtechnologies	Microplate and array readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com
Caliper Life Sciences	LabChip microfluidic systems for high-throughput screening	Hopkinton, Massachusetts	www.calipertech.com
Carl Zeiss Jena	plate::explorer and plate::screen systems for ultra-high-throughput screening	Göttingen, Germany	www.zeiss.com
Cellectricon	Patch-clamp equipment for ion-channel screening	Gothenburg, Sweden	www.cellectricon.se
Cellomics	ArrayScan HCS and KineticScan HCS systems for automated cell-based screening	Pittsburgh, Pennsylvania	www.cellomics.com
CyBio	Pipetting, liquid handling, incubation and imaging systems for automated screening	Jena, Germany	www.cybio-ag.com
Cytogration	Automated high-throughput cell culture systems for <i>in vitro</i> screening	Rockville, Maryland	www.cytogration.com
DiscoverRx	Cell-based assay development	Fremont, California	www.discoverx.com
Dynex Technologies	Microplate readers, washers and automated workstations	Chantilly, Virginia	www.dynextechnologies.com
EUGENEX Biotechnologies	Development of test cell lines	Taegerwilten, Switzerland	www.eugenex.com
GE Healthcare Bio-Sciences	Instruments, equipment and products for life-sciences research; LEADseeker for lead identification, IN Cell Analyzer confocal imaging system for rapid cellular assays	Little Chalfont, UK	www.amersham.co.uk
Jerini Array Technology	Substrate identification for kinases using pepSTAR peptide microarray technology	Berlin, Germany	www.jerini.com
KeyNeurotek	Functionality and tissue-based screening	Magdeburg, Germany	www.keyneurotek.de
Molecular Devices	Liquid-handling and microplate-processing equipment, imaging instruments and assay systems for integration into robotic systems	Sunnyvale, California	www.moleculardevices.com
OptiCell	Automated cell-culture devices	Westerville, Ohio	www.opticell.com
PerkinElmer Life Sciences	Automated systems for liquid handling and sample preparation for drug discovery and research; confocal laser scanning microscope for live-cell imaging	Boston, Massachusetts	lifesciences.perkinelmer.com
ProteDyne	BioCube automated systems	Windsor, Connecticut	www.proteDyne.com
QIAGEN	BioRobot multifunctional lab workstation and equipment for automated nucleic acid, protein and cell purification	Valencia, California	www.qiagen.com
RTS Life Sciences	Automated sample management, liquid handling, ultra-high-throughput screening	Manchester, UK	www.rts-group.com
SEPIAtec	HPLC/SPE equipment for high-throughput purification and fractionation	Berlin, Germany	www.sepiatec.com
Tecan	Laboratory automation, LabCD 'lab-on-a-disc' assays; Cellerity cell-culture system	Männedorf, Switzerland	www.tecan.com
TTP LabTech	Automated compound storage/retrieval systems	Royston, UK	www.ttplabtech.com
Velocity11	Automated high-throughput microplate processing, dispensers and pipetting stations	Palo Alto, California	www.velocity11.com
Vitro Biosciences	CellCards and CellPlex system for simultaneous assays of multiple cell types	Mountain View, California	www.vitrobio.com
Wave Biotech	Equipment for cell culture: Wave Bioreactor disposable cell-culture systems	Bridgewater, New Jersey	www.wavebiotech.com
Xcellsz	Immortalized cell lines, cell-based assays, screening	Newcastle upon Tyne, UK	www.xcellsz.com

General

Affymetrix	DNA microarrays; arrays; imaging systems and software	Santa Clara, California	www.affymetrix.com
Agilent	2100 Bioanalyzer for RNA, DNA and proteins; instruments for genomics and proteomics	Palo Alto, California	www.agilent.com
Analytikjena	Instrumentation for molecular spectroscopy, affinity measurements, and plate readers	Jena, Germany	www.analytikjena.com
Apogent Technologies	Labware and equipment for life sciences	Portsmouth, New Hampshire	www.apogent.com
Applied Biosystems	DNA and protein sequencers; mass spectrometers; chromatography and PCR analysis; peptide and nucleic-acid synthesizers; ICAT stable isotope labelling reagents	Foster City California	home.appliedbiosystems.com

Company	Products/activity	Location	URL
Applied Scientific Instrumentation	Automated systems for microscopy and fluorescence screening	Eugene, Oregon	www.asiimaging.com
BD Biosciences	Culture media, radioassay kits, FACS flow cytometers, monoclonal antibodies	Franklin Lakes, New Jersey	www.bd.com
Berthold Technologies	Bioanalytical instrumentation	Bad Wildbad, Germany	www.bertholdtech.com
Biacore	Analysis of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com
Bibby Sterilin	Laboratory glassware, consumables and equipment for life-sciences research	Stone, UK	www.bibby-sterilin.co.uk
Biophile	Ultra-low-temperature sample storage and retrieval system	Charlottesville, Virginia	www.biophileinc.com
Bio-Rad	Products, instruments and software for life-sciences research	Hercules, California	www.bio-rad.com
Biotage	High-performance liquid chromatography purification systems	Charlottesville, Virginia	www.biotage.com
BioTec	Liquid-handling lab instruments, pipetting workstation, dispenser, microplate washer	Tokyo, Japan	www.biotec.co.jp
Bioventures	Products and services for genomics and proteomics	Murfreesboro, Tennessee	www.bioventures.com
Brandel	Sample preparation and harvesting, superfusion systems and microdispensers	Gaithersburg, Maryland	www.brandel.com
Bruker	Automated platforms for mass spectrometry, NMR, EPR and MRI	The Woodlands, Texas	www.bruker.com
Chromagen	Fluorescence-based products, kits and services for protein detection and assay	San Diego, California	www.chromagen.com
Corning	Labware and laboratory equipment for life sciences	Corning, New York	www.corning.com
Deerac Fluidics	Nanovolume pipetting systems	Dublin, Ireland	www.allegro-technologies.com
EMD Biosciences	Novabiochem, Calbiochem, Novagen and Oncogene Research products	San Diego, California	splash.emdbiosciences.com
Eppendorf	Laboratory instrumentation and consumables for molecular and cell biology	Hamburg, Germany	www.eppendorf.com
Genetic Research Instrumentation	Equipment and consumables for molecular and cell-biology research	Braintree, UK	www.gri.co.uk
Genevac	Solvent evaporation for life-sciences research	Ipswich, UK	www.genevac.com
Gibco	Cell-culture media and consumables	Carlsbad, California	www.invitrogen.com
Gilson	Automated liquid-handling, pipetting and protein-crystallization instruments	Middleton, Wisconsin	www.gilson.com
Greiner Bio-One	Consumables for high-throughput research and diagnostics; microplates	Frickenhäusen, Germany	www.gbo.com/bioscience
Hamamatsu	Imaging and analysis systems	Hamamatsu City, Japan	www.hamamatsu.com
Hamilton Company	Microlab automated liquid-handling workstations	Reno, Nevada	hamiltoncomp.com
Hitachi High-Technologies	Automated instrumentation for mass spectroscopy, liquid chromatography, high-throughput purification, amino-acid analysis and spectroscopy	Tokyo, Japan	www.hitachi-hitec.com
IBM Life Sciences	DiscoveryLink platform for database integration; data-management systems	White Plains, New York	www-1.ibm.com/industries/lifesciences
Integra Biosciences	Equipment/consumables for sterilization, liquid handling, cell culture, sample storage	Baar, Switzerland	www.integra-biosciences.com
Invitrogen	Kits and reagents for cloning, genomics, proteomics, molecular and cell biology	Carlsbad, California	www.invitrogen.com
KBiosystems	Robotic instrumentation for high-throughput genomics and proteomics	Basilton, UK	www.kbiosystems.com
Kimble-Kontes	Laboratory glassware	Vineland, New Jersey	www.kimble-kontes.com
L.I.M.S.	StarLIMS laboratory information management system	South Hollywood, Florida	www.starlims.com
Labcyte	Pipetting workstations, microplate and magnetic bead washers, DNA extraction	Union City, California	www.labcyte.com
Medical Solutions	Laboratory diagnostics and tissue-analysis services	Nottingham, UK	www.medical-solutions.co.uk
Mitsui Knowledge Industry	LIMS	Tokyo, Japan	bio.mki.co.jp
MMI	Bioanalytical consumables, single-molecule counters and assay-automation facilities	Zurich, Switzerland	www.molecular-machines.com
Nalge Nunc International	Labware	Rochester, New York	www.nalgenunc.com
New England Biolabs	Reagents and kits for molecular biology	Beverly, Massachusetts	www.neb.com
Origene	Human cDNA clones; reagents and tools for genomics	Rockville, Maryland	www.origene.com
Pepscan	Chip with 1,200 protein kinase targets (peptides)	Lelystad, the Netherlands	www.pepscan.com
PhyNexus	Single-use pipette-tip-based chromatography columns for protein purification	San Jose, California	www.phynexus.com
Pierce Chemical	Components and consumables for automated systems in the life sciences	Rockford, Illinois	www.piercenet.com
PreAnalytiX	Products and reagents for nucleic-acid isolation	Hombrechtikon, Switzerland	www.preanalytix.com
Promega	Reagents and kits for genomics, proteomics and cellular analysis	Madison, Wisconsin	www.promega.com
Robbins Scientific	Labware, instruments and automated liquid-handling and dispensing equipment	Sunnyvale, California	www.robsci.com
RoboDesign	Automated equipment for biotechnology	Carlsbad, California	www.robodesign.com
Roche Diagnostics	Reagents and kits for molecular biology	Lewes, UK	www.roche-applied-science.com
Rohasys	Automated equipment for sample pretreatment, weighing, pH, conductivity, turbidity	Rijen, the Netherlands	www.rohasys.com
Sankyo Robotics	Multipurpose robots	Boca Raton, Florida	www.sankyo.com/robotics
Scinomix	Robotics and laboratory LIMS software for biotechnology	Earth City, Missouri	www.scinomix.com
Shimadzu Biotech	AXIMA MALDI QIT time-of-flight mass spectrometer for protein identification	Kyoto, Japan	www.shimadzu-biotech.net
Sirius	Instrumentation and analysis software for pKa, pH and membrane permeability	Forest Row, UK	www.sirius-analytical.com
Stratagene	Tools and reagents for molecular biology, genomics, proteomics, drug discovery	La Jolla, California	www.stratagene.com
Takara Bio	Reagents, kits and custom services for molecular biology, genomics, proteomics research; PCR enzymes; DNA microarrays; protein-refolding kit	Shiga, Japan	www.takara-bio.co.jp/english
TekCel	Low-temperature compound storage and retrieval	Hopkinton, Massachusetts	www.tekcel.com
The Automation Partnership	Automation for liquid handling, compound storage, high-throughput screening	Royston, UK	www.automationpartnership.com
Thermo Electron	Spectrometry, chromatography, microscopy, liquid handling, robotics	Waltham, Massachusetts	www.thermo.com
Tomtec	Liquid-handling workstations, solid-phase extraction workstations	Hamden, Connecticut	www.tomtec.com
Torcon Instruments	Precision fluid dispensing and automated instrumentation for microplate handling	Torrance, California	www.torconinstruments.com
Universal Technology	Automated storage and retrieval systems for the pharmaceutical industry	Pittsburgh, Pennsylvania	www.univtech.com
Upstate	KinaseProfiler specificity testing, IC50Profiler Express, crystallography assays, bulk protein production, chromatin immunoprecipitation kits, signalling proteins	Charlottesville, Virginia	www.upstate.com
Walden Precision Apparatus	pH meters, conductivity meters, colorimeters, spectrophotometers, oxygen meters	Cambridge, UK	www.wpaltd.co.uk
Warner Instruments	Electrophysiology and cell-biology equipment and products	Hamden, Connecticut	www.warneronline.com
Waters	Micromass instruments for mass spectroscopy; HPLC systems	Milford, Massachusetts	www.waters.com
Whatman	Separation technology, microplates, Biometra instruments	Maidstone, UK	www.whatman.com
Xiril	Pipetting robots	Hombrechtikon, Switzerland	www.xiril.com
Zinsser Analytic	Pipetting, extraction and liquid-handling robots	Frankfurt, Germany	www.zinsser-analytic.com

● see advertisement

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

The best years of our lives

I've lost count of how many times I've been told by tenured, senior researchers that doing their postdoc was the best part of their scientific life. I usually bite my tongue to stop myself from saying: "Then why don't you go back?" Senior researchers, now burdened with management and grant-writing duties, tend to remember their fellowship years as a period of pure scientific pursuit. And poverty is so much more romantic when you are no longer experiencing it.

But postdocs now are very different from the fellowships of old. Today's postdocs are feeling the squeeze. They often need to start seeking their first grant as soon as they begin their fellowship. And their long-term future is far from secure: there are far more fellows than there are tenure-track positions. Administrators, of course, prefer ever greater competition so that they can fill the few permanent slots with the best and the brightest — the scientific equivalent of rock stars.

Sadly, not every scientist can become a rock star; there simply isn't room on the stage. So what can the scientific equivalent of the opening acts and road crew do? Some simply leave — one researcher on a fixed-term contract at the University of Birmingham, UK, quit to train as a gas fitter — whereas others seek work outside the lab (see *EMBO Rep.* 5, 660–662; 2004).

The present system simply doesn't offer a solution. Too few tenure-track positions and a practice of short-term contracts, especially in Europe, mean that new scientists are constantly pushed into the system, but they often don't have a destination.

Clearly, there needs to be some middle ground. A mechanism that allows experienced scientists to continue doing benchwork, if they so choose — the kind of work that tenured management looks back upon so fondly.

Paul Smaglik
Naturejobs editor



Contents

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS